THE ANNALS
of
MATHEMATICAL STATISTICS

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Volume IX

1938

ᐧHE ANNALS
ᐧEMATICAL STATISTICS

EDITED BY

S. S. WILKS, *Editor*

IG J. NEYMAN

ᐧTH THE COÖPERATION OF

R. A. FISHER R. DE MISES
T. C. FRY E. S. PEARSON
H. HOTELLING H. L. RIETZ
W. A. SHEWHART

ᐧn in the ANNALS OF MATHEMATICAL STATISTICS
, Fine Hall, Princeton, New Jersey. Manuscripts
ᐧ-spaced with wide margins, and the original copy
ᐧtes should be reduced to a minimum and whenever
ᐧgraphy at the end of the paper; formulae in foot-
ᐧigures, charts, and diagrams should be drawn on
cloth in black India ink twice the size they are to
ᐧuested to keep in mind typographical difficulties
formulae.

ᐧceive only galley proofs. Fifty reprints without
Additional reprints and covers furnished at cost.

the ANNALS is $4.00 per year. Single copies $1.25.
ᐧt the following rates:

ᐧngle numbers $1.50.
Single numbers $1.25.

ᐧrders for back numbers and other business com-
ᐧ A. T. Craig, University of Iowa, Iowa City, Iowa.

ᐧTICAL STATISTICS is published quarterly by the
ᐧStatistics.

ᐧPOSED AND PRINTED AT THE
ᐧVAVERLY PRESS, INC.
ᐧALTIMORE, MD., U. S. A.

CONTENTS OF VOLUME IX

Four Dollars per annum

Back numbers available at the following prices:
Vols. I-IV $5 each. Single numbers $1.50
Vol. V to date $4 each. Single numbers $1.25

Made in United States of America

Address: ANNALS OF MATHEMATICAL STATISTICS
Post Office Box 171, Ann Arbor, Michigan

Office of the Institute of Mathematical Statistics:
Secretary: ALLEN T. CRAIG, University of Iowa
Iowa City, Iowa

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD.

COMBINED EXPANSIONS OF PRODUCTS OF SYMMETRIC POWER SUMS AND OF SUMS OF SYMMETRIC POWER PRODUCTS WITH APPLICATION TO SAMPLING[1]

BY PAUL S. DWYER

PREFACE

This article is divided into two parts. Part I has for its title "Combined Expansions of Products of Symmetric Power Sums and of Sums of Symmetric Power Products" and develops the general mathematical theory which is applied in Part II to "The Fundamentals of Sampling." Part II will appear in a latter issue of this journal.

Each part is treated as an organic unit and has its own introduction and bibliography. Each article is assigned a given number and each book is given a letter so that references can be indicated concisely in the body of the dissertation.

Each part is divided into chapters and sections. Braces are used to indicate the important formulas.

PART I. COMBINED EXPANSIONS OF PRODUCTS OF SYMMETRIC POWER SUMS AND OF SUMS OF SYMMETRIC POWER PRODUCTS

Introduction

The mathematical material which is presented here has proved useful in generalizing that portion of the fundamental theory of sampling in which relations are established between the moments of the sample and the moments of the parent population. It is the purpose to establish the theorems in algebraic form since they constitute an extension of partition and symmetric function theory and may be of value to someone not necessarily interested in sampling.

A great deal of work has been done in symmetric function theory but not much of this is of present value to the statistician. His problem deals with the "power sum" while the classical theory, for the most part, deals with the interrelations of elementary symmetric functions and monomial symmetric functions. Only one phase of the reasoning developed in this investigation seems to have received extensive consideration previously and that is the subject covered in Chapter III.

Previous authors have noted that much of symmetric function theory reduces, with a proper choice of notation, to partition theory. It is the plan of this treatise to present in Chapter I an outline of new partition theory which

[1] A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University of Michigan.

shows how the parts of one partition are combined to form the parts of another partition, and which serves as a means of expressing the main result of Chapters II, III, IV, V.

Chapter II shows how the formulas of Chapter I are applicable to the problem of finding products of power sums. The multiplication theorem for power sums, a generalization of the multinomial theorem, is stated in terms of power product sums and appropriate special cases are indicated.

Chapter III deals with the expansion of power product sums in terms of power sums and shows how the formulas of Chapter I may be used.

Chapter IV is the key chapter of the paper. The problem is to expand products of power sums in terms of power product sums, to multiply each power product sum by a quantity which is a uniquely defined function of the quantities composing the power product sum, and then to expand back in terms of all possible power sums. It is shown that the results can be written in a compact form which also utilizes the results of Chapter I. This result, as is shown in Part II, is directly applicable to the sampling problem of finding the moments of the sample moments in terms of the moments of the universe.

Extension is made to multivariate distributions in Chapter V.

Chapter I. The Combination of the Parts of Partitions

It is the purpose of this chapter to provide a precise notation which shows how the parts of one partition of r may be combined to form the parts of another partition of r. For example, 2111, a four part partition of 5, can be made into 32, a two part partition of 5, by combining the three unit parts into a new part or by combining the 2 with one of the unit parts to form the 3 and the other two unit parts to form the 2. This last formation can be made in three different ways since anyone of the unit parts might be combined with the 2. The combination of the parts of the partition 2111 to form the parts of the partition 32 is to be indicated symbolically by $P_{31} + 3P_{22}$ where the subscripts indicate the number of parts collected and the coefficients indicate the number of ways in which an equivalent collection can be made.

1. Definitions and Notation. a. *Partition* [G; 105] [K; I; 1] [16; 105]. We consider the integer r to be composed of r unit indistinguishable parcels and define the partitions of r to be all those different groupings into new parcels, each new parcel containing one or more unit parcels, such that each resultant grouping of parcels contains exactly all the original r unit parcels. For example the partitions of 4 are

4 ; 31 ; 22 ; 211 ; 1111

b. *Parts of Partitions.* The numbers of the grouped unit parcels indicate the parts of the partition. Thus the partitions of 4 above

4 ; 31 ; 22 ; 211 ; 1111 have respectively

1 ; 2 ; 2 ; 3 ; 4 parts.

The pattern 22 may also appear as 2^2. In general a ρ part partition of r is to be designated by

$p_1 p_2 p_3 \cdots p_\rho$ where the p's may or may not be equal and where $p_1 + p_2 + p_3 + \cdots + p_\rho = r$ or by

$$p_1^{\pi_1} \cdots p_s^{\pi_s} \text{ where } \begin{cases} p_1 \neq p_2 \neq p_3 \neq p_4 \neq \cdots \neq p_s \\ p_1\pi_1 + p_2\pi_2 + \cdots + p_s\pi_s = r \\ \pi_1 + \pi_2 + \cdots + \pi_s = \rho \end{cases}$$

c. *Order of Partitions.* When the parts of a partition are arranged in descending order we say that the partition is ordered. Thus

$$p_1 p_2 \cdots p_\rho \text{ is ordered if } p_1 \geq p_2 \geq p_3 \geq \cdots \geq p_\rho$$

and $$p_1^{\pi_1} \cdots p_s^{\pi_s} \text{ is ordered if } p_1 > p_2 > p_3 > \cdots > p_s.$$

For example, 21^2 is an ordered partition while 312 is not. Unless otherwise specified it is hereafter assumed that all partitions are ordered.

It is sometimes convenient to refer to the order of the partition which is the size of the largest part, p_1, when the partition is ordered. Thus the two part partition, 31, is of the order 3, while the four part partition, 1111, is of order 1. The set of the numbers $p_1^{\pi_1} \cdots p_s^{\pi_s}$ is to be known as the complete order.

These definitions of order and part are consistent with the usual definitions. [16; 105–106] [*K*; *I*; 1] [*G*; 100]. The concept of complete order, as far as I know, is not found in the literature.

d. *Weight of Partitions. Isobaric Partitions.* The weight of any partition is defined to be the sum of all the parts of the partition. Thus the weight of $p_1^{\pi_1} \cdots p_s^{\pi_s}$ is $p_1\pi_1 + p_2\pi_2 + \cdots + p_s\pi_s = r$. Partitions having the same weight are called isobaric. Thus 4 and 211 are isobaric partitions.

e. *Algebraic Partitions.* If the r original units are composed of a_1, a_2, a_3, $\cdots a_r$ nonseparable primary units, then the result of combining these in any possible way is to be called an algebraic partition since the r original units are now replaced by the r algebraic quantities $a_1, a_2, \cdots a_r$. Thus a_1, a_2, a_3 may be combined to form

$$a_1 + a_2 + a_3;\ \overline{a_1 + a_2}\cdot a_3;\ \overline{a_1 + a_3}\cdot a_2;\ \overline{a_2 + a_3}\cdot a_1;\ a_1\cdot a_2\cdot a_3$$

which are the algebraic partitions of $a_1 + a_2 + a_3$.

The parts of the algebraic partitions are the resulting combinations while the order and complete order, which indicate the numbers of algebraic expressions combined, agree with the order and complete order of the partitions in which the a's are unity. The weight, which is equal to the sum of the parts, is indicated by $w = a_1 + a_2 + \cdots + a_r$. Thus if $a_1 = 5$, $a_2 = 4$, and $a_3 = 3$, $w = 12$. It is to be noted that the algebraic partitions are formed by combining the parts 5, 4, 3 and not by combining all parts of 12.

Now $a_1\ a_2\ \cdots\ a_r$ is itself a partition of weight $w = a_1 + a_2 + \cdots + a_r$. If groups of the a's are alike it may be written

$$a_1^{\alpha_1}\, a_2^{\alpha_2}\ \cdots\ a_h^{\alpha_h} \text{ where}$$

$$a_1\alpha_1 + a_2\alpha_2 + \cdots + a_h\alpha_h = w$$

$$\alpha_1 + \alpha_2 + \cdots + \alpha_h = r.$$

Algebraic partitions having the same weight are called isobaric.

f. *Partition Combination Notation.* Let $\begin{pmatrix} 1^r \\ p_1^{\pi_1} \cdots p_s^{\pi_s} \end{pmatrix}$ indicate the number of different ways the r units, ordinary or algebraic, can be collected to form the partition. Thus $\begin{pmatrix} 1^5 \\ 32 \end{pmatrix}$ indicates the number of ways in which the five units can be collected to form a partition with three units in one part and two units in the other. Since the three units forming the first part can be selected in ${}_5C_3$ ways and since this selection automatically indicates the other two units forming the second part, it follows that

$$\begin{pmatrix} 1^5 \\ 32 \end{pmatrix} = {}_5C_3 = {}_5C_2 = 10. \text{ It is to be noted that } \begin{pmatrix} 1^4 \\ 22 \end{pmatrix} = 3 \neq {}_4C_2 = 6$$

for if the four unit parts are a_1, a_2, a_3, a_4, then the three 22 partitions are

$$\overline{a_1 + a_2}\cdot\overline{a_3 + a_4};\ \overline{a_1 + a_3}\cdot\overline{a_2 + a_4};\ \overline{a_1 + a_4}\cdot\overline{a_2 + a_3}$$

since

$$\overline{a_3 + a_4}\cdot\overline{a_1 + a_2};\ \overline{a_2 + a_4}\cdot\overline{a_1 + a_3};\ \overline{a_2 + a_3}\cdot\overline{a_1 + a_4}$$

are essentially the same groupings as the first three indicated.

2. Formula for $\begin{pmatrix} 1^r \\ p_1^{\pi_1} \cdots p_s^{\pi_s} \end{pmatrix}$. In establishing this formula we first take the case in which no part is repeated. i.e. $\pi_1 = \pi_2 \cdots \pi_s = 1$ and $p_1 + p_2 + p_3 \cdots + p_s = r$. In this case the formula becomes

$$\begin{pmatrix} 1^r \\ p_1 p_2 \cdots p_s \end{pmatrix} = \frac{r!}{p_1!\, p_2! \cdots p_s!}$$

This results from the fact that the p_1 units can be grouped in ${}_rC_{p_1}$ different ways. The p_2 units then in ${}_{r-p_1}C_{p_2}$ different ways, the p_3 units then in ${}_{r-p_1-p_2}C_{p_3}$ different ways etc. So that

$$\begin{pmatrix} 1^r \\ p_1 p_2 \cdots p_s \end{pmatrix} = {}_rC_{p_1}\cdot{}_{r-p_1}C_{p_2}\cdot{}_{r-p_1-p_2}C_{p_3} \cdots {}_{r-p_1-p_2\cdots p_{s-1}}C_{p_s}$$

$$= \frac{r!}{p_1!p_2! \cdots p_s!} \qquad \text{Compare [B; 49][19; 12]}$$

If however $p_1 = p_2 = \cdots = p_s$, then the same partition has been used $s!$ different times since $p_1, p_2, \cdots, p_s$ may be interchanged in $s!$ different ways, so that

$$\binom{1^r}{p_1^s} = \frac{r!}{(p_1!)^s s!}$$

By similar reasoning

$$\binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} = \frac{r!}{(p_1!)^{\pi_1} (p_2!)^{\pi_2} \cdots (p_s!)^{\pi_s} \pi_1! \pi_2! \cdots \pi_s!} \qquad \{1\}$$

Compare [19; 12, 13] [I; II, 252]

3. **Values of** $\binom{a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_h^{\alpha_h}}{p_1^{\pi_1} p_2^{\pi_2} \cdots p_s^{\pi_s}}$. The number of ways in which the r parts of $a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_h^{\alpha_h}$ may be collected to form the ρ parts of $p_1^{\pi_1} \cdots p_s^{\pi_s}$ may be indicated by

$$\binom{a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_h^{\alpha_h}}{p_1^{\pi_1} p_2^{\pi_2} \cdots p_s^{\pi_s}}. \quad \text{Thus} \quad \binom{2111}{32} = 4 \quad \text{and} \quad \binom{1111}{22} = 3$$

Formulas useful is evaluating this expression can be worked out from the results of this paper. A table of values of this expression for $w \leq 8$ has been given by the author [19; 29–32].

4. **Notation for Combining the Parts of a Partition. Table I.** We wish to indicate not only the number of ways in which a given r part partition of weight w can be grouped to form a ρ part partition of weight w, but also the number of parts of the r part partition grouped to form each of the ρ parts of the ρ part partition. As indicated in the opening paragraph, $P_{31} + 3P_{22} = P\binom{2111}{32}$ serves this purpose for the case in which the parts 2111 are collected to form 32. $P\binom{a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_h^{\alpha_h}}{p_1^{\pi_1} p_2^{\pi_2} \cdots p_s^{\pi_s}}$ serves this purpose in the more general case. Its expansion gives sums of P functions whose subscripts are the numbers of parts combined and whose coefficients are the number of ways of forming the partitions from the parts. For example

$$P\binom{31}{4} = P_1; \qquad P\binom{111}{21} = 3P_{21}; \text{ etc.}$$

The use of $P\binom{a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_h^{\alpha_h}}{p_1^{\pi_1} p_2^{\pi_2} \cdots p_s^{\pi_s}}$ is so fundamental to the present approach that a table is provided showing the different values when $w \leq 6$.

Table I gives values of the function when $w = 1, 2, 3, 4, 5, 6$. The values $a_1^{\alpha_1} \cdots a_h^{\alpha_h}$ are given in the left hand columns and the values of $p_1^{\pi_1} \cdots p_s^{\pi_s}$ in the top row. The partitions are ordered from the top and from the left. To

TABLE I

Values of $P\begin{pmatrix} a_1^{\pi_1} a_2^{\pi_2} \cdots a_r^{\pi_r} \\ p_1^{\pi_1} p_2^{\pi_2} \cdots p_s^{\pi_s} \end{pmatrix}$ when $W \leqq 6$.

$W = 1$

$a \backslash p$	1
1	P_1

$W = 2$

$a \backslash p$	2	11
2	P_1	
11	P_2	P_{11}

$W = 6$

$a \backslash p$	6	51	42	33	411	321	222	31^3	2^21^2	21^4	1^6
6	P_1										
51	P_2	P_{11}									
42	P_2		P_{11}								
33	P_2			P_{11}							
411	P_3	$2P_{21}$	P_{12}		P_{111}						
321	P_3	P_{21}	P_{21}	P_{12}		P_{111}					
222	P_3		$3P_{21}$				P_{111}				
31^3	P_4	$3P_{31}$	$3P_{22}$	P_{13}	$3P_{211}$	$3P_{121}$		P_{1111}			
2^21^2	P_4	$2P_{31}$	$2P_{31}$ P_{22}	$2P_{22}$	P_{211}	$4P_{211}$	P_{112}		P_{1111}		
21^4	P_5	$4P_{41}$	P_{41} $6P_{32}$	$4P_{23}$	$6P_{311}$	$4P_{311}$ $12P_{221}$	$3P_{122}$	$4P_{2111}$	$6P_{1211}$	P_{11111}	
1^6	P_6	$6P_{51}$	$15P_{42}$	$10P_{33}$	$15P_{411}$	$60P_{321}$	$15P_{222}$	$20P_{3111}$	$45P_{2211}$	$15P_{21111}$	P_{111111}

$W = 3$

$a \backslash p$	3	21	111
3	P_1		
21	P_2	P_{11}	
111	P_3	$3P_{21}$	P_{111}

$W = 4$

$a \backslash p$	4	31	22	211	1111
4	P_1				
31	P_2	P_{11}			
22	P_2		P_{11}		
211	P_3	$2P_{21}$	P_{12}	P_{111}	
1111	P_4	$4P_{31}$	$3P_{22}$	$6P_{211}$	P_{1111}

$W = 5$

$a \backslash p$	5	41	32	311	221	2111	11111
5	P_1						
41	P_2	P_{11}					
32	P_2		P_{11}				
311	P_3	$2P_{21}$	P_{12}	P_{111}			
221	P_3	P_{21}	$2P_{21}$		P_{111}		
2111	P_4	$3P_{31}$	P_{31} $3P_{22}$	$3P_{211}$	$3P_{121}$	P_{1111}	
11111	P_5	$5P_{41}$	$10P_{32}$	$10P_{311}$	$15P_{221}$	$10P_{2111}$	P_{11111}

find a given value, say $P\binom{2111}{22}$ we note that $w = 5$, look for 2111 on the left and 32 at the top. The result is $P_{31} + 3P_{22}$. In the table the order of the subscripts is important in indicating the number of parts collected to form the respective parts of the ordered $p_1^{\pi_1} \cdots p_s^{\pi_s}$.

The values in the table previously mentioned [19; 29–32] may be obtained when $w \leqq 6$ by placing every P in Table I equal to unity.

5. Value of $P(a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_h^{\alpha_h})$. The parts of the partition $a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_h^{\alpha_h}$ may be collected to form a large number of partitions of the type $p_1^{\pi_1} \cdots p_s^{\pi_s}$. Thus the parts of the partition 2111 may be collected to form 5, 41, 32, 311, 221, 2111. We denote by $P(2111)$ the values

$$P\binom{2111}{5}5 + P\binom{2111}{41}41 + P\binom{2111}{32}32 + P\binom{2111}{311}311 + P\binom{2111}{221}221$$
$$+ P\binom{2111}{2111}2111 = P_4 5 + 3P_{31}41 + [P_{31} + 3P_{22}]\,32 + 3P_{211}311 + 3P_{211}221 + P_{1111}2111$$

and in general

$$P(a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_h^{\alpha_h}) = \sum P\binom{a_1^{\alpha_1} \cdots a_h^{\alpha_h}}{p_1^{\pi_1} \cdots p_s^{\pi_s}} p_1^{\pi_1} \cdots p_s^{\pi_s} \quad \ldots\ldots\ldots \{2\}$$

where the summation holds for every partition $p_1^{\pi_1} \cdots p_s^{\pi_s}$ which can be formed by combining parts of $a_1^{\alpha_1} \cdots a_h^{\alpha_h}$. The values of $P(a_1^{\alpha_1} \cdots a_h^{\alpha_h})$ for $w \leqq 6$ are given in the rows of Table I. Thus the value of $P(2111)$ above is found along the row 2111 where $w = 5$.

6. Values of $P(1^r)$ and $P(a^r)$. When $a_1 = 1$ and $\alpha_1 = r$ we have

$$P(1^r) = \sum P\binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} p_1^{\pi_1} \cdots p_s^{\pi_s}$$

and since there are $\binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ different ways of forming $p_1^{\pi_1} \cdots p_s^{\pi_s}$ from the r units and each way is indicated by $P_{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ we have

$$P(1^r) = \sum \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} p_1^{\pi_1} \cdots p_s^{\pi_s}. \qquad \{3\}$$

When $r = 2, 3, 4$, etc., we get

$$P(1^2) = P_2\,2 + P_{11}\,1^2$$
$$P(1^3) = P_3\,3 + 3P_{21}\,21 + P_{111}\,1^3$$
$$P(1^4) = P_4\,4 + 4P_{31}\,31 + 3P_{22}\,22 + 6P_{211}\,211 + P_{1111}\,1^4$$

etc.

as indicated in Table I.

Similarly when $a_1 = a$ and $\alpha_1 = r$ {2} becomes

$$P(a^r) = \sum \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} (ap_1)^{\pi_1} (ap_2)^{\pi_2} \cdots (ap_s)^{\pi_s} \qquad \{3'\}$$

since there are $\binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ different ways of forming the partition $(ap_1)^{\pi_1} (ap_2)^{\pi_2} \cdots (ap_s)^{\pi_s}$ from the r equal a's and each way is indicated by $P_{p_1^{\pi_1} \cdots p_s^{\pi_s}}$.
For example

$$P(a) = P\binom{a}{a} a = P_1\, a$$

$$P(a^2) = P\binom{a^2}{2a} + P\binom{a^2}{aa} = P_2\, 2a + P_{11}\, a^2$$

$$P(a^3) = P\binom{a^3}{3a} + P\binom{a}{2a \cdot a} + P\binom{a}{a \cdot a \cdot a} = P_3\, 3a + 3P_{21}\, 2a \cdot a + P_{111}\, a^3$$

$$P(a^4) = P_4\, 4a + 4P_{31}\, 3a \cdot a + 3P_{22}\, 2a \cdot 2a + 6P_{211}\, 2a \cdot a^2 + P_{1111}\, a^4.$$

7. **Values of $P(a_1 a_2 \cdots a_r)$.** From the definition

$$P(a_1) = P\binom{a_1}{a_1} a_1 = P_1\, a_1$$

$$P(a_1 a_2) = P\binom{a_1 a_2}{a_1 + a_2} + P\binom{a_1 a_2}{a_1 a_2}$$

$$= P_2\, \overline{a_1 + a_2} + P_{11}\, a_1\, a_2$$

$$P(a_1 a_2 a_3) = P\binom{a_1 a_2 a_3}{a_1 + a_3 + a_3}(a_1 + a_2 + a_3) + P\binom{a_1 a_2 a_3}{\overline{a_1 + a_2} \cdot a_3}(\overline{a_1 + a_2} \cdot a_3)$$

$$+ P\binom{a_1 a_2 a_3}{\overline{a_1 + a_3} \cdot a_2}(\overline{a_1 + a_3} \cdot a_2) + P\binom{a_1 a_2 a_3}{\overline{a_2 + a_3} \cdot a_1}(\overline{a_2 + a_3} \cdot a_1)$$

$$+ P\binom{a_1 a_2 a_3}{a_1 a_2 a_3}(a_1 a_2 a_3) = P_3(a_1 + a_2 + a_3)$$

$$+ P_{21}\{(\overline{a_1 + a_2} \cdot a_3) + (\overline{a_1 + a_3} \cdot a_2) + (\overline{a_2 + a_3} \cdot a_1)\} + P_{111}(a_1 a_2 a_3)$$

etc.

Now if complete order of the general partition indicates the number of a's collected to form the partition, the subscripts of the P's are the respective complete orders. If we indicate the sum of partitions having the same complete order by the term "partition type" and indicate the partition type composed of all terms having the same complete order

$$p_1^{\pi_1} \cdots p_s^{\pi_s} \text{ by } T_{p_1^{\pi_1} \cdots p_s^{\pi_s}}$$

then

$$
\begin{aligned}
P(a_1) &= P_1 T_1 \\
P(a_1 a_2) &= P_2 T_2 + P_{11} T_{11} \\
P(a_1 a_2 a_3) &= P_3 T_3 + P_{21} T_{21} + P_{111} T_{111} \\
P(a_1 a_2 a_3 a_4) &= P_4 T_4 + P_{31} T_{31} + P_{22} T_{22} + P_{211} T_{211} + P_{1111} T_{1111}
\end{aligned}
$$

etc.

and in general

$$P(a_1 a_2 \cdots a_r) = \sum P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} T_{p_1^{\pi_1} \cdots p_s^{\pi_s}} \tag{4}$$

This formula can be used in writing the formula of Table II or formulas of weight greater than 6. Thus

$$P(543) \text{ is given by } P_3 T_3 + P_{21} T_{21} + P_{111} T_{111}$$

where

$$T_3 = 12,\ T_{21} = 9 \cdot 3 + 8 \cdot 4 + 7 \cdot 5, \text{ and } T_{111} = 5 \cdot 4 \cdot 3$$

where the dots do not indicate multiplication, but merely the separation of the parts.

In general $T_{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ is composed of $\binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ partitions since this is the number of ways in which the a's can be combined to form partitions having the same complete order, $p_1^{\pi_1} \cdots p_s^{\pi_s}$.

Formula {3′} is a special case of this formula. If the a's are all equal, the $\binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ partitions are equal so that $T_{p_1^{\pi_1} \cdots p_s^{\pi_s}} = \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} (ap_1)^{\pi_1} \cdots (ap_s)^{\pi_s}$. Substitution in {4} gives {3′}. Similarly {4} gives {3} when all the a's are unity.

8. **Generalization from Symmetry.** The function $P(a_1 a_2 \cdots a_r)$ is a symmetric function of the parts $a_1, a_2, \cdots a_r$, i.e., the interchange of any two of the parts does not change the value of the function. It is possible to use this fact as a basis of generalization and to derive {4} from {3} by its use. From {3} we have

$$P(1^r) = \sum \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} p_1^{\pi_1} \cdots p_s^{\pi_s} \tag{3}$$

where $\binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ is the number of the equivalent partitions which can be formed from the r units. In case the r units are replaced by the r different a's, there will result $\binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ different partitions having the same complete

order. These $\binom{1^r}{p_1^{\pi_1}\cdots p_s^{\pi_s}}$ different partitions defined by $T_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$ replace the $\binom{1^r}{p_1^{\pi_1}\cdots p_s^{\pi_s}}$ equivalent partitions of {2} and we have

$$P(a_1a_2\cdots a_r) = \sum P_{p_1^{\pi_1}\cdots p_s^{\pi_s}}T_{p_1^{\pi_1}\cdots p_s^{\pi_s}} \tag{4}$$

9. **The Recursion Rule.** It is possible to establish a recursion property by which the value of $P(a_1a_2\cdots a_ra_{r+1})$ can be obtained from the value of $P(a_1a_2\cdots a_r)$. We note, from the results of Table I or by {4} that

$$P(3) = P_1 3$$

$$P(32) = P_2 5 + P_{11}32$$

$$P(321) = P_3 6 + P_{21}51 + P_{21}42 + P_{12}33 + P_{111}111$$

$P(32)$ is obtained from $P(3)$ by symbolic multiplication of its expansion, $P_1(3)$, by the expansion of $P(2)$, $P_1(2)$. This symbolic multiplication is accomplished by adding the 2 to the 3 and also suffixing the 2 to the 3. If the 2 is added, the subscripts of the P's are added while if suffixed, the subscripts of the P's are suffixed.

More generally if $P(a_1) = P_1(a_1)$ and $P(a_2) = P_1(a_2)$, then the result $P(a_1a_2) = P_2(\overline{a_1+a_2}) + P_{11}(a_1)\,(a_2)$ is obtained by multiplying $P_1(a_1)$ by $P_1(a_2)$ [or $P_1(a_2)$ by $P_1(a_1)$] symbolically if the subscripts are added when the a's are added and suffixed when the a's are suffixed. Similarly $P(a_1a_2) = P_2(\overline{a_1+a_2}) + P_{11}(a_1)\,(a_2)$ when multiplied by $P(a_3) = P_1(a_3)$ gives

$$P(a_1a_2a_3) = P_3(\overline{a_1+a_2+a_3}) + P_{21}(\overline{a_1+a_2}\cdot a_3) + P_{21}(\overline{a_1+a_3}\cdot a_2) + P_{12}(a_1\cdot\overline{a_2+a_3}) + P_{111}(a_1a_2a_3)$$

when the rule of multiplication is the adding of a_3 in turn to every part of every partition with the appropriate adding of subscripts and the suffixing of a_3 to every partition with the corresponding suffixing of subscripts. It is important to note that the P coefficient of $a_1\cdot\overline{a_2+a_3}$ is P_{12} and not P_{21} although the term could be written $P_{21}\overline{a_2+a_3}\cdot a_1$. The applications do not demand the retention of a given order of subscripts though the continued application of the recursion rule does demand it.

In general the value of $P(a_1\cdots a_ra_{r+1})$ can be obtained from the value of $P(a_1a_2\cdots a_r)$ by the symbolic multiplication of the expansion $P(a_1a_2\cdots a_r)$ by $P_1(a_{r+1})$ since all possible algebraic partitions of $a_1+a_2+\cdots+a_r+a_{r+1}$ are obtained from all possible algebraic partitions of $a_1+a_2+\cdots+a_r$ by adding a_{r+1} in turn to each part of each partition and by suffixing it to each partition. The corresponding P subscript, indicating the number of a's collected, is increased by 1.

The recursion rule is useful in checking the entries of Table I. As a matter

of fact Table I was computed with its use and the order of the subscripts is that which results from its use. The rule is also useful in finding values when $w > 6$. For example, since

$$P(321) = P_3 6 + P_{21} 51 + P_{21} 42 + P_{12} 33 + P_{111} 321$$

$$\begin{aligned} P(3221) &= P_4 8 + P_{31} 62 + P_{31} 71 + P_{22} 53 + P_{211} 512 + P_{31} 62 + P_{22} 44 \\ &+ P_{211} 422 + P_{22} 53 + P_{13} 35 + P_{121} 332 + P_{211} 521 + P_{121} 341 + P_{112} 323 \\ &+ P_{1111} 3212 = P_4 8 + P_{31} 71 + 2\, P_{31} 62 + (P_{31} + 2P_{22}) 53 + P_{22} 44 \\ &+ 2P_{211} 521 + P_{211} 431 + P_{211} 422 + 2P_{211} 332 + P_{1111} 3221. \end{aligned}$$

A useful check is based on the fact that the sum of the P coefficients of $P(a_1 \cdots a_r)$ should equal the sum of the coefficients of $P(1^r)$. In the above illustration the sum of the coefficients is $P_4 + 4P_{31} + 3P_{22} + 6P_{211} + P_{1111}$ as desired.

10. **Use of the P Function Formulas.** The P function formulas, as defined, represent concisely the ways in which the parts of a given partition may be combined to get the parts of other partitions. They are also useful in writing expansions of certain partition functions whose expanded values are expressed in terms of other partition functions. They are used, in this paper, in expressing the multinomial theorem, the multiplication theorem for power sums, the expansions of power product sums in terms of power sums, expansions of monomial symmetric functions in terms of power sums, the double expansion theorem itself, the coefficients in the double expansion theorem as well as the sampling laws of Part II. They are also useful in representing the expansions of different moment functions and can be associated with important concepts of mathematics and statistics such as, for example, the differences of 0. Such applications, however, are not pertinent to the line of reasoning which is developed in Chapters II, III, IV, V.

Chapter II

It is the purpose of this chapter to obtain formulas for the expansion of power sums.

11. **Definitions.** a. *Power Sum.* Let x be a variable which is restricted to the N variates, $x_1, x_2, x_3, \cdots, x_N$. Then the a-th power sum of the variable indicated by (a) is defined to be

$$(a) = x_1^a + x_2^a + \cdots + a_N^a = \sum_{i=1}^{N} x_i^a \tag{5}$$

It is assumed for the purposes of this paper that a is a positive integer or 0.

b. *Power Product Sum.* The expression $\sum_{i \neq j} x_i^{a_1} x_j^{a_2}$ is to be called a power

product sum since it is composed of the sum of products of the powers of the variates. It is to be denoted by $(a_1 \cdot a_2)$ or $(a_1 a_2)$. Thus $\sum_{i \neq j} x_i^3 x_j^2 = (3 \cdot 2)$ or (32). The value $(a \cdot a) = (a^2) = \sum_{i \neq j} x_i^a x_j^a$ is a special case of $(a_1 a_2)$ where $a_2 = a_1 = a$. In general the power product sum is defined by the right hand member and indicated by the left hand member of

$$(a_1 a_2 \cdots a_r) = \sum_{i_1 \neq i_2 \neq i_3 \neq \cdots \neq i_r} x_{i_1}^{a_1} x_{i_2}^{a_2} \cdots x_{i_p}^{a_p} \tag{6}$$

If $i_1 = i_2$, the power product sum becomes

$$(\overline{a_1 + a_2} \cdot a_3 \cdot a_4 \cdots a_r) = \sum_{i_1 = i_2 \neq i_3 \neq i_4 \neq \cdots \neq i_r} x_{i_1}^{a_1} x_{i_2}^{a_2} \cdots x_{i_p}^{a_p} \tag{7}$$

There are many different definitions since there are many different ways of indicating equality relations among the i's. Each results in a unique power product sum which is to be called, for brevity, a power product. If the a's are all unity, there are many duplicates. Thus for the grouping $p_1^{\pi_1} \cdots p_s^{\pi_s}$, there are $\begin{pmatrix} 1^r \\ p_1^{\pi_1} \cdots p_s^{\pi_s} \end{pmatrix}$ equal power products $(p_1^{\pi_1} \cdots p_s^{\pi_s})$. In the more general case we can let $T_{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ represent the $\begin{pmatrix} 1^r \\ p_1^{\pi_1} \cdots p_s^{\pi_s} \end{pmatrix}$ different power products having the same complete order, $p_1^{\pi_1} \cdots p_s^{\pi_s}$. We may represent any one of these forms having this complete order by

$$(q_1 q_2 q_3 \cdots q_\rho)$$

where $\quad q_1 + q_2 + q_3 + \cdots + q_\rho = w$

or by $\quad (q_1^{x_1} q_2^{x_2} \cdots q_t^{x_t})$*

where $\quad q_1 x_1 + q_2 x_2 + \cdots + q_t x_t = w$

and $\quad x_1 + x_2 + \cdots + x_t = \rho.$

c. *Symmetric Functions.* Both the power sum and the power product are symmetric functions of the variates since the interchange of any x_i with any x_j does not change the value of the function. Also the powder product having ρ parts is composed of $N^{(\rho)}$ products of powers since the first group of equal i's may be selected in N ways, the next group in $N - 1$ ways etc.

d. *Monomial Symmetric Function.* It is customary to use the monomial symmetric function which is defined as

$$\sum_{i_1 < i_2 < i_3 < \cdots < i_\rho} x_{i_1}^{q_1} x_{i_2}^{q_2} \cdots x_{i_\rho}^{q_\rho}$$

* "It was intended that the letter representing the exponents of the q's should be the Greek 'chi,' and not the English 'x.'"

and which we designate by $M(q_1 \cdots q_\rho)$ or by $M(q_1^{x_1} \cdots q_t^{x_t})$. This function is not useful for our purposes since the number of terms in its expansion varies with the number of repeated q's. For example if $N = 3$ and $q_1 \gtrless q_2$; $M(q_1 q_2)$ $= x_1^{q_1}x_2^{q_2} + x_1^{q_1}x_3^{q_2} + x_2^{q_1}x_3^{q_2} + x_2^{q_1}x_1^{q_2} + x_3^{q_1}x_1^{q_2} + x_3^{q_1}x_2^{q_2} = (q_1 q_2)$ while if $q_1 = q_2 = q$

$$M(q^2) = x_1^q x_2^q + x_1^q x_3^q + x_2^q x_3^q = \frac{(q^2)}{2!}.$$

The monomial symmetric function keeps the number of product terms a minimum by eliminating all repeated terms while the power product sum keeps the number of product terms the same by the use of repeated terms, when some of the parts are alike.

12. The Formula Connecting $(q_1^{x_1} q_2^{x_2} \cdots q_t^{x_t})$ and $M(q_1^{x_1} \cdots q_t^{x_t})$. The power product is composed of $N^{(\rho)}$ products, each of which is repeated $x_1!x_2!$ $\cdots x_t!$ times. The monomial symmetric function is composed of the $\dfrac{N^{(\rho)}}{x_1!x_2! \cdots x_t!}$ different products which, when repeated $x_1!x_2! \cdots x_t!$ times, gives the $N^{(\rho)}$ terms above. Hence

$$(q_1^{x_1} \cdots q_t^{x_t}) = x_1!x_2! \cdots x_t!\, M(q_1^{x_1} \cdots q_t^{x_t}) \tag{8}$$

$$M(q_1^{x_1} \cdots q_t^{x_t}) = \frac{1}{x_1!x_2! \cdots x_t!}(q_1^{x_1} \cdots q_t^{x_t}) \tag{9}$$

In the special case in which $q_1 = 1$ and $x_1 = \rho$

$$(1^\rho) = \rho!\, M(1^\rho) \quad \text{and} \quad M(1^\rho) = \frac{(1^\rho)}{\rho!} \tag{10}$$

The function, $M(1^\rho)$ is commonly called an elementary symmetric function. We refer to the corresponding (1^ρ) as the unitary power product sum.

13. Correspondence of Partitions and Power Products. To each power product $(q_1^{x_1} \cdots q_t^{x_t})$ there corresponds an algebraic partition $q_1^{x_1} \cdots q_t^{x_t}$ having ρ parts and weight $w = a_1 + a_2 + \cdots + a_r$.

This follows at once from the definitions and notation. Thus if $w = a_1 + a_2 + a_3$, the power product

$$\sum_{i_1 = i_2 \neq i_3} x_{i_1}^{a_1} x_{i_2}^{a_2} x_{i_3}^{a_3} = \sum_{i_1 \neq i_2} x_{i_1}^{a_1+a_2} x_{i_2}^{a_3} = (\overline{a_1 + a_2} \cdot a_3)$$

is, by notation, associated with the partition $\overline{a_1 + a_2} \cdot a_3$. Conversely each algebraic partition, when enclosed in parentheses, represents a power product sum.

This proposition is useful in that it enables one to establish a relationship between the theory of power product sums and the theory of partitions. Earlier writers have used a similar correspondence in relating the theory of monomial

symmetric functions to that of partitions. See for instance [3; 106], [4; 5] [5; I; 7].

Due to this correspondence we do not hesitate to apply such terms as part, order, complete order, similar, etc. to the power product as well as to the algebraic partition. Also the sum of all power products $(q_1^{\pi_1} \cdots q_t^{\pi_t})$ having the same complete order is represented by $T(p_1^{\pi_1} \cdots p_s^{\pi_s})$. This represents the sum of $\binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ similar power products.

14. **The Multiplication Theorem for Power Sums.** The correspondence property enables us to derive a theorem, to be known as the multiplication theorem, which expresses products of power sums in terms of power products. The type of argument is introduced by establishing simple cases of the theorem

$$(a_1)(a_2) = \Big(\sum_{i=1} x_i^{a_1}\Big)\Big(\sum_{i=1} x_i^{a_2}\Big) = \sum_{\substack{i_1=1\\ i_2=1}}^{N} x_{i_1}^{a_1} x_{i_2}^{a_2}$$

$$= \sum_{i_1=i_2} x_{i_1}^{a_1} x_{i_2}^{a_2} + \sum_{i_1\neq i_2} x_1^{a_1} x_2^{a_2} = (a_1 + a_2) + (a_1 a_2)$$

$$(a_1)(a_2)(a_3) = \left(\sum x_i^{a_1}\right)\left(\sum x_i^{a_2}\right)\left(\sum x_i^{a_3}\right) = \sum_{i_1,i_2,i_3} x_{i_1}^{a_1} x_{i_2}^{a_2} x_{i_3}^{a_3}$$

$$= (a_1 + a_2 + a_3) + (\overline{a_1 + a_2}\cdot a_3) + (\overline{a_1 + a_3}\cdot a_2) + (a_1\cdot\overline{a_2 + a_3}) + (a_1\cdot a_2\cdot a_3)$$

since the value $\sum_{i_1,i_2,i_3}$ is broken into

$$\sum_{i_1=i_2=i_3}, \quad \sum_{i_1=i_2\neq i_3}, \quad \sum_{i_1=i_3\neq i_2}, \quad \sum_{i_1\neq i_2=i_3}, \quad \sum_{i_1\neq i_2\neq i_3}.$$

In general, when $r \leqq N$

$$(a_1)(a_2) \cdots (a_r) = \sum_{i_1,i_2,\cdots i_r} x_{i_1}^{a_1} x_{i_2}^{a_2} \cdots x_{i_r}^{a_r}$$

and this can be broken into summations featuring different equality relations. These summations define all the different power product sums of weight $w = a_1 + a_2 + \cdots + a_r$. The different algebraic partitions of $a_1 + a_2 + a_3 + \cdots + a_r$ correspond to the different power product sums. It follows at once that the value of $(a_1)(a_2) \cdots (a_r)$ is obtained by writing each algebraic partition of $a_1 + a_2 + \cdots + a_r$, enclosing it in parentheses to represent a power product, and adding. More symbolically we have

$$(a_1)(a_2) \cdots (a_r) = \sum (q_1^{\pi_1} \cdots q_t^{\pi_t}) \tag{11}$$

where $q_1^{\pi_1} \cdots q_t^{\pi_t}$ represents any algebraic partition of $a_1 + a_2 + \cdots + a_r$. and the summation holds for all such partitions or by

$$(a_1)(a_2) \cdots (a_r) = \sum T(p_1^{\pi_1} \cdots p_s^{\pi_s}) \tag{12}$$

where $T(p_1^{\pi_1} \cdots p_s^{\pi_s})$ represents the $\binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ similar power products and the summation holds for each different complete order.

For example $(a_1)(a_2)(a_3) = T(3) + T(21) + T(111)$ and $T(3) = (a_1 + a_2 + a_3)$, $T(21) = (\overline{a_1 + a_2 \cdot a_3}) + (\overline{a_1 + a_3 \cdot a_2}) + (\overline{a_2 + a_3 \cdot a_1})$, and $T(111) = (a_1 \cdot a_2 \cdot a_3)$.

The theorem has been established on the assumption that $r \leqq N$. If such is not the case it is possible to satisfy the assumption by adding additional variates, $x_{N+1}, x_{N+2}, \cdots x_r$, all 0, without changing the value of the power sums or of the product of the power sums since the added terms are always 0. Thus

$$(x_1^a + x_2^a)(x_1^b + x_2^b)(x_1^c + x_2^c) = (x_1^a + x_2^a + x_3^a)(x_1^b + x_2^b + x_3^b)(x_1^c + x_2^c + x_3^c)$$

when $x_3 = 0$

Then

$$(a)(b)(c) = (a + b + c) + (\overline{a + b \cdot c}) + (\overline{a + c \cdot b}) + (\overline{b + c \cdot a}) + (a \cdot b \cdot c)$$

which is

$$\sum x_i^{a+b+c} + \sum_{i \neq j} x_i^{a+b} x_j^c + \sum_{i \neq j} x_i^{a+c} x_j^b + \sum_{i \neq j} x_i^{b+c} x_j^a + \sum_{i \neq j \neq k} x_i^a x_j^b x_k^c$$

The term $\sum_{i \neq j \neq k} x_i^a x_j^b x_k^c = 0$ since every product composing it contains an $x_3 = 0$. The other power product sums are to be applied to the original variates only since the terms involving x_3 are 0 in every case.

In general, if $r > N$, it is only necessary to write out the power product sums having N or less parts since all those having more than N parts will be 0.

15. The Multiplication Theorem Using the Results of Chapter I. Comparison of {12} with {4} shows that {12} can be obtained from {4} by placing $P(a_1 \cdots a_r) = (a_1)(a_2) \cdots (a_r)$, $T_{p_1^{\pi_1} \cdots p_s^{\pi_s}} = T(p_1^{\pi_1} \cdots p_s^{\pi_s})$ and $P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} = 1$. Since this can be done for all values of a and r it follows at once that the entire theory of Chapter I is applicable to the present problem. For example Table I shows that

$$P(321) = P_3 6 + P_{21} 51 + P_{21} 42 + P_{12} 33 + P_{111} 321$$

and it follows that

$$(3)(2)(1) = (6) + (51) + (42) + (33) + (321)$$

It should be noted that it is possible to use the table previously published [19; 29–32] since the entries in this table are the values obtained when $P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} = 1$. The value $(3)(2)(1)$ may also be checked from this table.

16. The Multinomial Theorem. The multinomial theorem is a special case of the multiplication theorem for power sums in which the power sums are all equal. If $a_1 = a_2 = \cdots = a_r = 1$,

$$T(p_1^{\pi_1} \cdots p_s^{\pi_s}) = \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} (p_1^{\pi_1} \cdots p_s^{\pi_s})$$

and {12} becomes

$$(1)^r = \sum \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} (p_1^{\pi_1} \cdots p_s^{\pi_s}) \tag*{\{13\}}$$

which is the multinomial theorem in terms of power product sums. Special cases are

$$(1)^2 = (2) + (11)$$
$$(1)^3 = (3) + 3(21) + (111)$$
$$(1)^4 = (4) + 4(31) + 3(22) + 6(211) + (1111)$$

etc.

The result of {13} may also be obtained immediately from {3} by placing $P(1^r) = (1)^r$, $P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} = 1$, and $p_1^{\pi_1} \cdots p_s^{\pi_s} = (p_1^{\pi_1} \cdots p_s^{\pi_s})$.

A more general form of the multinomial theorem is that in which $a_1 = a_2 = \cdots = a_r = a$. In this case

$$(x_1^a + x_2^a + \cdots + x_N^a)^r = (\mathrm{a})^r$$

and {12} gives

$$(a)^r = \sum \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} ((ap_1)^{\pi_1} \cdots (ap_s)^{\pi_s}) \tag*{\{14\}}$$

where $((ap_1)^{\pi_1} \cdots (ap_s)^{\pi_s})$ has parts $ap_1, \cdots, ap_s$. Thus

$$(a)^3 = (3a) + 3(2a \cdot a) + (a^3)$$

so that

$$(2)^3 = (6) + 3(4 \cdot 2) + (2^3).$$

The result {14} may also be obtained immediately from {3′} by placing $P(a^r) = (a)^r$, $P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} = 1$, and $(ap_1)^{\pi_1} \cdots (ap_s)^{\pi_s} = ((ap_1)^{\pi_1} \cdots (ap_s)^{\pi_s})$. When $N = 2$, {13} gives the binomial theorem

$$(1)^r = \sum \binom{1^r}{p_1 p_2} (p_1 p_2)$$

special cases of which are

$$(1)^2 = (2) + (11)$$
$$(1)^3 = (3) + 3(21)$$
$$(1)^4 = (4) + 4(31) + 3(22)$$
$$(1)^5 = (5) + 5(41) + 10(32)$$

etc.

These can be readily translated to the usual form. Thus

$$(a+b)^4 = a^4 + b^4 + 4(a^3b + b^3a^3) + 3(a^2b^2 + b^2a^2).$$

In a similar manner the trinomial theorem appears as

$$(1)^r = \sum \binom{1^r}{p_1 p_2 p_3} (p_1 p_2 p_3)$$

A special case of the multinomial theorem {13} is also useful in writing N^r in terms of sums of $N^{(\rho)}$. When the variates are all unity the power sums are all N, and the power product sums are the number of terms in the partition representing it. If a partition has ρ parts the number of terms in it is $N^{(\rho)}$. We then have

$$N^r = \sum \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} N^{(\rho)} \qquad \{15\}$$

Special cases are

$$N^2 = N + N^{(2)}$$

$$N^3 = N + 3N^{(2)} + N^{(3)}$$

$$N^4 = N + 4N^{(2)} + 3N^{(2)} + 6N^{(3)} + N^{(4)} = N + 7N^{(2)} + 6N^{(3)} + N^{(4)}$$

etc.

17. **The Use of Monomial Symmetric Functions.** It is possible to express the results in terms of the monomial symmetric functions by means of {8}. Thus

$$\begin{aligned}(2)(2)(2) &= (6) + 3(42) + (222) \\ &= M(6) + 3M(42) + 6M(222).\end{aligned}$$

In general, Table I may be used to express products of power sums in terms of monomial symmetric functions. It is only necessary to place every $P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} = 1$ and to multiply by the factorials indicating the repeated entries at the head of each column. The table [19; 29–32] may be used similarly.

The multinomial theorem in terms of monomial symmetric functions becomes

$$(1)^r = \sum \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} \pi_1!\pi_2! \cdots \pi_s! M(p_1^{\pi_1} \cdots p_s^{\pi_s})$$

and by {1}

$$(1)^r = \sum \frac{r!}{(p_1!)^{\pi_1}(p_2!)^{\pi_2} \cdots (p_s!)^{\pi_s}} M(p_1^{\pi_1} \cdots p_s^{\pi_s}) \qquad \{16\}$$

as it is conventionally stated.

18. **The Multiplication Theorem from the Multinomial Theorem.** It is possible to use generalization from symmetry in deriving the multiplication theorem from the multinomial theorem though this can not well be done from its conventional statement (16). The monomial symmetric function does not have the property that $M(a \cdot b) = M(a \cdot a)$ when $b = a$ while $(a \cdot b)$ does become $(a \cdot a)$ when $b = a$. The first step then is to reduce $\{16\}$ to power product sums by means of $\{9\}$. We then have

$$(1)^r = \sum \frac{r!}{(p_1!)^{\pi_1}(p_2!)^{\pi_2} \cdots (p_s!)^{\pi_s} \pi_1!\pi_2! \cdots \pi_s!}(p_1^{\pi_1} \cdots p_s^{\pi_s})$$

Next it is necessary to introduce the factor $\begin{pmatrix} 1^r \\ p_1^{\pi_1} \cdots p_s^{\pi_s} \end{pmatrix}$ for there are many equal terms for each value $p_1^{\pi_1} \cdots p_s^{\pi_s}$ when the a's are all unity. This is very easy in this case since the value of the coefficient of $(p_1^{\pi_1} \cdots p_s^{\pi_s})$ is $\begin{pmatrix} 1^r \\ p_1^{\pi_1} \cdots p_s^{\pi_s} \end{pmatrix}$. It follows at once that

$$(1)^r = \sum \begin{pmatrix} 1^r \\ p_1^{\pi_1} \cdots p_s^{\pi_s} \end{pmatrix}(p_1^{\pi_1} \cdots p_s^{\pi_s}).$$

Suppose that the r units are replaced by $a_1 a_2 \cdots a_r$. Then the $\begin{pmatrix} 1^r \\ p_1^{\pi_1} \cdots p_s^{\pi_s} \end{pmatrix}$ power products, $(p_1^{\pi_1} \cdots p_s^{\pi_s})$ will be replaced by the $\begin{pmatrix} 1^r \\ p_1^{\pi_1} \cdots p_s^{\pi_s} \end{pmatrix}$ different power products composing $T(p_1^{\pi_1} \cdots p_s^{\pi_s})$. It follows at once that

$$(a_1)(a_2) \cdots (a_r) = \sum T(p_1^{\pi_1} \cdots p_s^{\pi_s}).$$

19. **The Determination of the Coefficient of a Given Power Product in the Expansion of a Product of Power Sums.** In some cases we wish to determine the coefficient of a given power product without computing the complete expansion. This is given by $P\begin{pmatrix} a_1^{\alpha_1} \cdots a_h^{\alpha_h} \\ q_1^{x_1} \cdots q_t^{x_t} \end{pmatrix}$ where the P coefficients are unity. Thus the coefficient of (32) in the expansion of (2)(1)(1)(1) is found from $P\begin{pmatrix} 2111 \\ 32 \end{pmatrix} = P_{31} + 3P_{22}$ and is 4.

20. **Relation to Previous Results.** The multiplication theorem may be viewed as a generalization of the multinomial theorem. A more general proof, applicable to multivariate problems, could be presented with the use of more involved notation. It seems wise rather to present the simpler one variate case and to emphasize the principle of generalization from symmetry which will enable us to write the multivariate laws with relative ease.

The general problem discussed here seems to have received a very small

amount of consideration as much of the extensive classical theory of symmetric functions is limited to the interrelations of the elementary symmetric functions and the monomial symmetric functions.

A monumental work on symmetric functions not subject to this limitation is the Combinatory Analysis of MacMahon [K]. MacMahon provided a technique for multiplying power sums in many variables as a special case of a more general theory. [K; II, 321].

Some of the work on alternants is closely related to the problem of products of power sums although the alternant, as usually defined, is limited to the case in which $r = N$ [I; II, 446]. For an example the reader is referred to a development by Muir [L; 335–6].

Thiele (1889) gave tables[2] of products of power sums in terms of monomial symmetric functions for partition products of weight $\leqq 8$ [H; 114–117]. J. R. Roe has later given one for $w \leqq 10$ [N; Plates 17, 18]. Statisticians have sometimes stated the results in nontabular form. See for example, the multiplication formulas of Church [13; 81–83] [14; 370–1], whose results may not at first appear to agree with those above since Church has used a less compact notation and, of course, the monomial symmetric function.

The chief contributions of the present attack are

1. The use of the formulas and tables of Chapter I in writing expansions of products of power sums.

2. The use of power product sums in place of monomial symmetric functions which makes feasible.

3. Generalization from symmetry.

Chapter III

It is the purpose of this chapter to establish formulas giving the expansion of power products in terms of products of power sums.

21. The Binet (Waring) Identities. It is customary to introduce this subject with formulas for $M(a \cdot b)$, $M(a \cdot b \cdot c)$, etc. so we first derive the formulas for $(a \cdot b)$, $(a \cdot b \cdot c)$, etc. We may use the results of Chapter II since the problem here is the inverse of the multiplication problem. By the multiplication theorem

$$(a)(b) = (a + b) + (a \cdot b)$$

$$(a)(b)(c) = (a + b + c) + (\overline{a + b} \cdot c) + (\overline{a + c} \cdot b) + (\overline{b + c} \cdot a) + (a \cdot b \cdot c)$$

$$(a + b)(c) = (a + b + c) + (\overline{a + b} \cdot c)$$

[2] These tables are not accessible to me, but Thiele refers to them in his "Theory of Observations."

so we get

$$(a \cdot b) = (a)(b) - (a + b) \tag{17}$$

$$(a \cdot b \cdot c) = (a)(b)(c) - (a + b)(c) - (a + c)(b) - (b + c)(a) + 2(a + b + c) \tag{18}$$

Similarly

$$\begin{aligned}(a \cdot b \cdot c \cdot d) = {} & (a)(b)(c)(d) - (a + b)(c)(d) - (a + c)(b)(d) \\ & - (a + d)(b)(c) - (b + c)(a)(d) - (b + d)(a)(c) \\ & - (c + d)(a)(b) - (a + b)(c + a) - (a + c)(b + d) \\ & - (a + d)(b + c) + 2(a + b + c)(d) + 2(a + b + d)(c) \\ & + 2(a + c + d)(b) + 2(b + c + d)(a) - 6(a + b + c + d)\end{aligned} \tag{19}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \{20\}$$

When $a \neq b \neq c \neq d$, {18}, {19}, {20} are also the formulas for $M(ab)$, $M(abc)$, $M(abcd)$. These formulas are quite commonly attributed to Binet who gave them in 1812 in connection with certain proofs of determinant theory [1; 284] [I; I; 81]. Waring should be given credit (see Miscellanea Analytica 1762). Binet gave no proof. The reader is also referred to the earlier work of Paoli [A; section 28].

A much more adequate treatment was given by Hirsch in the early 19th century [B; 35–38]. He wrote, out the terms for $M(a \cdot b \cdot c \cdot d \cdot e)$ and indicated a scheme for extending the results. More than this he proved that any "numerical expression"—his term for monomial symmetric function—can be reduced to numerical expression having one less part [B; 26]. The continued application of this theorem leads eventually to numerical expressions having only one part, i.e. to power sums. Hence all numerical expressions can be reduced to power sums [B; 27, 32].

Recent authors give essentially the same proof. See for example Bocher [J; 241–242] who states the theorem, "Every symmetric polynomial is a linear combination with constant coefficients of a certain number of the Σ's." See also O'Toole [16; 114] and Burnside and Panton [E; 167]. Thus modern authors provide a proof of the fact that $M(a_1 \cdots a_r)$ can be expanded in terms of power sums but most of them fail to provide a formula giving this precise expansion. Even MacMahon after writing the values of $M(\lambda\mu)$, $M(\lambda\mu r)$, $M(\lambda^2)$, $M(\lambda^3)$ avoids the immediate generalization by stating [K; I; 7], "In actual practice there are easier ways of calculating the many part functions and the general formula is of little importance."

While MacMahon's statement has a certain amount of truth in that any given monomial symmetric function may be computed from others having one less part by the recursion property described by Hirsch, yet there are many cases in which a definite formula, rather than a method, is desirable. A formula

particularly is demanded by the statistician who is working with a large number of monomial symmetric functions simultaneously. See for example the remarks and efforts of Carver [15; 103–104, 119–120], Church [14; 373, 377–378], and O'Toole [16; 115].

Some authors have provided solutions and it appears that statisticians are not entirely familiar with all the work which has previously been done. It is the aim of the remainder of this chapter to suggest references which make previous work available to statisticians as well as to present a logical and quite complete development. The main results are not essentially new although their explicit statement in the language of power products is necessary for the development of the next chapter. The argument features the easy generalization from symmetry. The value of (1^r) is expressed in such a form that the value $(a_1 \cdots a_r)$ may be obtained immediately from it.

22. The Value of (1^r) from Waring's Expansion for the Elementary Symmetric Function. We first derive the formula (1^r) from the conventional Waring's expression for p_m in terms of the power sums. Burnside and Panton [E; II; 92] give this as

$$p_m = \sum \frac{(-1)^{r_1+r_2+\cdots+r_m} s_1^{r_1} s_2^{r_2} \cdots s_m^{r_m}}{\Gamma(r_1+1)\,\Gamma(r_2+1)\cdots\Gamma(r_m+1)\,2^{r_2}3^{r_3}\cdots m^{r_m}} \qquad \{20\}$$

where $p_m = (-1)^m M(1^m)$ and where $S_1^{r_1} \cdots S_m^{r_m}$ is any $r_1 + r_2 + \cdots + r_m$ part partition of m. When $m = r$ and $p_1^{\pi_1} \cdots p_s^{\pi_s}$ is any ρ part partition of m, {20} becomes

$$(-1)^r M(1^r) = \sum \frac{(-1)^\rho (p_1)^{\pi_1} \cdots (p_s)^{\pi_s}}{p_1!^{\pi_1} \cdots p_s!^{\pi_s}\, \pi_1!\,\pi_2! \cdots \pi_s!} \qquad \{21\}$$

Dividing by $(-1)^r$ and noting that $(-1)^{\rho-r} = (-1)^{r-\rho}$ we have

$$M(1^r) = \sum \frac{(-1)^{r-\rho} (p_1)^{\pi_1} \cdots (p_s)^{\pi_s}}{p_1!^{\pi_1} \cdots p_s!^{\pi_s}\, \pi_1!\,\pi_2! \cdots \pi_s!} \qquad \{22\}$$

and hence that

$$(1^r) = r!M(1^r) = \sum \frac{(-1)^{r-\rho} r!\, (p_1)^{\pi_1} \cdots (p_s)^{\pi_s}}{p_1!^{\pi_1} \cdots p_s!^{\pi_s}\, \pi_1!\,\pi_2! \cdots \pi_s!} \qquad \{23\}$$

A second proof of {23}, given in the next sections, does not assume the formula {20} and develops by easy stages. Although somewhat longer than the method above, it contacts much of the work that has been done in this field. It also provides two useful arithmetic checks dealing with the coefficients which the more analytic method above does not provide. Those who are familar with {20} above and are interested in the immediate development of the argument with the use of {23} should turn to the equivalent {38} of section 28.

23. The Newtonian Formulas. The development begins with the well known formulas connecting the power sums and the elementary symmetric functions which appeared in Newton's Arithmetica Universalis. These formulas are given by Bocher (J; 244) as follows

$$S_k - p_1 S_{k-1} + \cdots + (-1)^{k-1} p_{k-1} S_1 + (-1)^k k p_k = 0 \quad k = 1, 2, \cdots \qquad \{24\}$$

where S_k is the sum of the k-th powers and p is the i-th elementary symmetric function.

So many proofs of this theorem are accessible that a repetition here is hardly justifiable. A proof using calculus was given by Bocher (J; 243). Proofs using algebra only were given by Hirsch (B; 16) and Chrystal (F; I, 437). Muirhead (9; 66–70) gave three proofs of which the second is perhaps best adapted to the present development.

24. The Determinant Equivalent of (1^r). It is usual to solve the Newtonian equations for the power sums (J; 244) but our objective is the solution in terms of the power sums. The equations are

$$\begin{aligned}
&p_1 &&= (1)\\
&(1)p_1 - 2p_2 &&= (2)\\
&(2)p_2 - (1)p_2 + 3p_3 &&= (3)\\
&(3)p_1 - (2)p_2 + (1)p_3 - 4p_4 &&= (4)\\
&\cdots\cdots\cdots\cdots
\end{aligned}$$

whence

$$p_r = \frac{\begin{vmatrix}
1 & 0 & 0 & \cdots\cdots & 0 & (1)\\
(1) & -2 & 0 & \cdots\cdots & 0 & (2)\\
(2) & -(1) & 3 & \cdots\cdots & 0 & (3)\\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\
(r-2) & -(r-3) & (r-4) & \cdots\cdots & (-1)^{r-2}(r-1) & (r-1)\\
(r-1) & -(r-2) & (r-3) & \cdots\cdots & (-1)^{r-2}(1) & (r)
\end{vmatrix}}{\begin{vmatrix}
1 & 0 & 0 & \cdots\cdots & 0 & 0\\
(1) & -2 & 0 & \cdots\cdots & 0 & 0\\
(2) & -(1) & 3 & \cdots\cdots & 0 & 0\\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\
(r-2) & -(r-3) & (r-4) & \cdots\cdots & (-1)^{r-2}(r-1) & 0\\
(r-1) & -(r-2) & (r-3) & \cdots\cdots & (-1)^{r-2}(1) & (-1)^{r-1}r
\end{vmatrix}}$$

Next, factor out all the negative signs in the even numbered columns in each determinant. The number of these columns of negative signs is the same as

the number in the denominator if r is odd. If r is even, there is one more in the denominator. Hence the negative signs may be dropped in both determinants if $(-1)^{r-1}$ is inserted in the numerator. Furthermore the value of the determinant in the denominator is $r!$ Next, change the numerator by moving the r-th column to the first column position and inserting the compensating factor $(-1)^{r-1}$. If Δ_r represents the resulting numerator determinant, the value of p_r becomes

$$p_r = \frac{(-1)^{r-1}(-1)^{r-1}}{r!}\Delta_r$$

and

$$\Delta_r = r!p_r = (1^r).$$

We have then

$$(1^r) = \Delta_r = \begin{vmatrix} (1) & 1 & 0 & \cdots\cdots & 0 & 0 \\ (2) & (1) & 2 & \cdots\cdots & 0 & 0 \\ (3) & (2) & (1) & \cdots\cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots\cdots & \cdot & \cdot \\ (r-1) & (r-2) & (r-3) & \cdots\cdots & (1) & r-1 \\ (r) & (r-1) & (r-2) & \cdots\cdots & (2) & (1) \end{vmatrix} \qquad \{25\}$$

The determinant has received the attention of earlier writers {19; 3}. Generalizations of it will be mentioned at the close of the chapter. Its expansion in terms of power sums is known and may be written

$$\Delta_r = \sum \frac{(-1)^{r-\rho} r!\,(p_1)^{\pi_1} \cdots (p_s)^{\pi_s}}{p_1!^{\pi_1} \cdots p_s!^{\pi_s}\,\pi_1!\pi_2! \cdots \pi_s!} \qquad \{26\}$$

where
$$p_1\pi_1 + p_2\pi_2 + \cdots + p_s\pi_s = r$$

and
$$\pi_1 + \pi_2 + \cdots + \pi_s = \rho.$$

See for example O'Toole (16; 113).

It is at once evident that {26} is equivalent to {23}. Those who are familiar with the expansion of Δ_r above may wish to turn immediately to {38} of section 28 since the intervening sections are devoted to a rather detailed and rigorous expansion of the determinant. This development follows, in a general way, that given by Mola (5; 190–195).

25. **The Expansion of $(1^r) = \Delta_r$.** The determinant, Δ_r, is a special type of determinant which is known as a recurrent. There is a simple recursion property which is useful in its expansions in terms of products of power sums.

$$\Delta_{r+1} = \begin{vmatrix} (1) & 1 & 0 & \cdots\cdots & 0 & 0 \\ (2) & (1) & 2 & \cdots\cdots & 0 & 0 \\ (3) & (2) & (1) & \cdots\cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ (r) & (r-1) & (r-2) & & (1) & r \\ (r+1) & (r) & (r-1) & & (2) & (1) \end{vmatrix}$$

If we expand Δ_{r+1} in terms of the $(r+1)$st column we have

$$\Delta_{r+1} = (1)\Delta_r - r\overline{\Delta}_r \tag{27}$$

where $\overline{\Delta}_r$ represents the determinant Δ_r with every power sum in the r-th row increased by unity. It is only necessary to arrive at some method of designating these terms if the above recurrence formula is to be applied. This can be done by inserting the power sum (1) before the other power sums which it is to multiply. Also in forming $\overline{\Delta}_r$ add unity to the first power sums in the expansion of Δ_r being careful to retain the previous order. Thus

$$\left.\begin{aligned}
\Delta_1 &= (1) \\
\Delta_2 &= (1)(1) - 1(1+1) = (1)(1) - (2) \\
\Delta_3 &= (1)[(1)(1) - (2)] - 2[(2)(1) - (3)] = (1)(1)(1) - (1)(2) \\
&\quad - 2(2)(1) + 2(3) \\
\Delta_4 &= (1)(1)(1)(1) - (1)(1)(2) - 2(1)(2)(1) + 2(1)(3) - 3(2)(1)(1) \\
&\quad + 3(2)(2) + 6(3)(1) - 6(4) \\
\Delta_5 &= (1)(1)(1)(1)(1) - (1)(1)(1)(2) - 2(1)(1)(2)(1) + 2(1)(1)(3) \\
&\quad - 3(1)(2)(1)(1) + 3(1)(2)(2) + 6(1)(3)(1) - 6(1)(4) \\
&\quad - 4(2)(1)(1)(1) + 4(2)(1)(2) + 8(2)(2)(1) - 8(2)(3) \\
&\quad + 12(3)(1)(1) - 12(3)(2) - 24(4)(1) + 24(5) \\
&\quad \text{etc.}
\end{aligned}\right\} \tag{28}$$

By collection of repeated terms and recalling that $(1^r) = \Delta_r$, the expansion becomes

$$\left.\begin{aligned}
(1) &= (1) \\
(1^2) &= (1)^2 - (2) \\
(1^3) &= (1)^3 - 3(2)(1) + 2(3) \\
(1^4) &= (1)^4 - 6(2)(1)^2 + 3(2)^2 + 8(3)(1) - 6(4) \\
(1^5) &= (1)^5 - 10(2)(1)^3 + 15(2)(2)(1) + 20(3)(1)(1) - 20(3)(2) \\
&\quad - 30(4)(1) + 24(5). \\
&\cdots\cdots\cdots
\end{aligned}\right\} \tag{29}$$

It is possible to write values of (1^r) in terms of power sums though the practical difficulty increases as r increases. Also continued use of the recursion formula {27} is apt to lead to error. Two simple checks are available. If D_r represents the sum of the coefficients of the expansion of (1^r) and $| D_r |$ represents the sum of the absolute values of these coefficients, then

$$D_r = 0 \qquad \text{when } r > 1 \tag{30}$$

$$| D_r | = r! \tag{31}$$

The proof of {30} and {31} follows directly from {27} since the coefficients of Δ_r and $\mathbf{\Delta}_r$ are the same. Thus $D_{r+1} = (1 - r)D_r$ and $| D_{r+1}| = (1 + r)| D_r |$. Since $D_2 = 0$ it follows that $D_3, D_4, \cdots, D_r = 0$ and since $| D_2 | = 2!$ it follows that $| D_3 |, | D_4 |, \cdots, | D_r |$ are $3!, 4!, r!$ respectively.

26. Determination of the Coefficient of Any Ordered Product of Power Sums in the Expansion of Δ_r. We next attempt to revise the process outlined above so as to get the formulas {29} without going through the work of writing out {28}. We note first that every product of power sums in the expansion of (1^r) in {28} has been obtained from (1) by a succession of $r - 1$ operations which were either prefixes (when the (1) was prefixed) or raises (when the (1) was added). Also the order of the power sums in a given term indicates which operations have been prefixes and which raises. For example (1)(1)(1)(1)(1) results from 4 prefixes while (5) results from 4 raises. The term (3)(2) results from 1 raise, 1 prefix, and 2 raises respectively, while the term (2)(3) results from 2 raises, a prefix, and a raise. The product $(p_4)(p_3)(p_2)(p_1)$ results from prefixes when $r = p_1$, $r = p_1 + p_2$, $r = p_1 + p_2 + p_3$ and raises at all other times.

The sign of the coefficient of $(p_4)(p_3)(p_2)(p_1)$ can be determined when we recall that each raise is accompanied by a multiplication by $-r$ while each prefix is accompanied by no change in the coefficient. There have been $p_1 - 1 + p_2 - 1 + p_3 - 1 + p_4 - 1 = p_1 + p_2 + p_3 + p_4 - 4$ raises so the sign is $(-1)^{r-\rho}$ where $p_1 + p_2 + p_3 + p_4 = r$. More generally if $(p_\rho) \cdots (p_3)(p_2)(p_1)$ is a term in the expansion of (1^r) where $p_\rho + \cdots + p_3 + p_2 + p_1 = r$ the number of changes in the sign is $p_1 - 1 + p_2 - 1 + \cdots + p_\rho - 1 = p_1 + p_2 + \cdots + p_\rho - p = r - \rho$. It follows at once that those products of power sums in the expansion of (1^r) which have the same number of factors, ρ, also have the same sign and that this sign is $(-1)^{r-\rho}$.

In determining the numerical part of the coefficient we note that each prefix is accompanied by a multiplication by unity which can be written in the form $\frac{r}{r}$. Each raise is accompanied by a multiplication by r so there appears in the numerator the product of all possible values of r and in the denominator the product of those values of r corresponding to each prefix. For example the numerical coefficient of $(p_4)(p_3)(p_2)(p_1)$ is

$$\frac{(p_4 + p_3 + p_2 + p_1 - 1)!}{(p_3 + p_2 + p_1)(p_2 + p_1)(p_1)} = \frac{(p_4 + p_3 + p_2 + p_1)!}{(p_4 + p_3 + p_2 + p_1)(p_3 + p_2 + p_1)(p_2 + p_1)(p_1)}$$

Similarly the coefficient, without sign of $(p_\rho)(p_{\rho-1}), \cdots (p_3)(p_2)(p_1)$ in the expansion of (1^r) is

$$\frac{(p_\rho + p_{\rho-1} + \cdots + p_3 + p_2 + p_1)!}{(p_\rho + p_{\rho-1} + \cdots + p_3 + p_2 + p_1)(p_{\rho-1} + \cdots + p_3 + p_2 + p_1) \cdots (p_3 + p_2 + p_1)(p_2 + p_1)(p_1)} \qquad \{32\}$$

The denominator of {32} has a certain resemblance to a factorial. Thus $4! = (1 + 1 + 1 + 1)\,(1 + 1 + 1)\,(1 + 1)\,(1)$ in which the successive factors are found by dropping the first unit. The corresponding algebraic expression $(p_4 + p_3 + p_2 + p_1)\,(p_3 + p_2 + p_1)\,(p_2 + p_1)\,(p_1)$ is found in the same way and might be called an "algebraic factorial." It might be designated by

$$(p_4 + p_3 + p_2 + p_1)¡$$

It should be noted that the order of the terms in the algebraic factorial is significant. Thus $(p_2 + p_1)¡ \neq (p_1 + p_2)¡$ unless $p_1 = p_2$.

The coefficient of $(p_\rho)(p_{\rho-1}) \cdots (p_2)(p_1)$ in the expansion of (1^r) may now be written

$$(-1)^{r-\rho} \frac{r!}{(p_\rho + \cdots + p_2 + p_1)¡} \qquad \{33\}$$

For example the coefficient

$$(2)(1)(2) \text{ is } (-1)^{5-3} \frac{5!}{5 \cdot 3 \cdot 2} = 4$$

$$(1)(2)(2) \text{ is } (-1)^{5-3} \frac{5!}{5 \cdot 4 \cdot 2} = 3$$

$$(2)(2)(1) \text{ is } (-1)^{5-3} \frac{5!}{5 \cdot 3 \cdot 1} = 8$$

and the total coefficient of all terms involving (2)(2)(1) is 15.

With a less formal notation we might designate the sum of the $\rho!$ "algebraic factorials" which can be formed from $p_\rho, p_{\rho-1}, \cdots p_2, p_1$ by

$$\sum (p_\rho + p_{\rho-1} + \cdots + p_2 + p_1)¡$$

and the sum of their reciprocals by

$$\sum \frac{1}{(p_\rho + p_{\rho-1} + \cdots + p_2 + p_1)¡}$$

This notation calls for the inclusion of all the $\rho!$ algebraic factorials even though some of them may be alike. If

$$\pi_1, \pi_2, \cdots, \pi_s$$

indicate the numbers of repeated p's

$$\sum \frac{1}{(p_\rho + p_{\rho-1} + \cdots + p_2 + p_1)¡} = \pi_1!\pi_2! \cdots \pi_s! \sum{}' \frac{1}{(p_\rho + p_{\rho-1} + \cdots + p_2 + p_1)¡} \qquad \{34\}$$

where $\sum'$ holds for the $\dfrac{r!}{\pi_1!\pi_2!\cdots\pi_s!}$ non-repeated terms.

In general the total coefficient of $(p_1)(p_2) \cdots (p_\rho)$ in the expansion of (1′) is obtained by adding all possible terms {33} in which the same p's occur in different positions in the product. Every possible different position grouping of the p's is present but once since it is dependent solely on the unique order in which prefixes and raises have been combined to produce that particular position grouping. The number of these position groupings varies with the number of repeated p's. The sum of the coefficients of these position groupings of the same p's, i.e. the total coefficient of $(p_1)(p_2) \cdots (p_\rho)$ is then given by

$$(-1)^{r-\rho} r! \sum{}' \frac{1}{(p_\rho + \cdots + p_1)¡}$$

which can be written by means of {34}

$$(-1)^{r-\rho} \frac{r!}{\pi_1!\pi_2!\cdots\pi_s!} \sum \frac{1}{(p_\rho + \cdots + p_1)¡}.$$

The formula for (1) may be written

$$(1') = \sum (-1)^{r-\rho} \frac{r!}{\pi_1!\pi_2!\cdots\pi_s!} \sum \frac{1}{(p_\rho + \cdots + p_1)¡} (p_1)(p_2) \cdots (p_\rho) \qquad \{35\}$$

27. **Theorem on Algebraic Factorials.** The result {35} can be further simplified by the theorem

$$\sum \frac{1}{(p_\rho + \cdots + p_2 + p_1)¡} = \frac{1}{p_\rho p_{\rho-1} \cdots p_2 p_1} \qquad \{36\}$$

which is proved by mathematical induction.

A. It is true when $\rho = 2$, since

$$\sum \frac{1}{(p_2 + p_1)¡} = \frac{1}{(p_2 + p_1)¡} + \frac{1}{(p_1 + p_2)¡} = \frac{1}{p_2 + p_1}\left[\frac{1}{p_1} + \frac{1}{p_2}\right] = \frac{1}{p_1 p_2}$$

B. If it is true for $\rho = k$, it is true for $\rho = k + 1$ since

$$\sum \frac{1}{(p_{k+1} + p_k + \cdots + p_2 + p_1)¡} = \frac{1}{p_{k+1} + \cdots + p_2 + p_1}\left[\sum \frac{1}{(p_k + \cdots + p_1)¡} + \sum_{p_{k+1}} \sum \frac{1}{(p_k + p_{k-1} + \cdots + p_1)¡}\right]$$

where $\sum_{p_{k+1}}$ gives the k terms in which p_{k+1} replaces $p_k, p_{k-1}, \cdots, p_2, p_1$ respectively. Now if {36} is true when $\rho = k$

$$\sum \frac{1}{(p_{k+1} + \cdots + p_1)_1^1} = \frac{1}{p_{k+1} + \cdots + p_1}\left[\frac{p_{k+1} + p_k + p_{k-1} + \cdots + p_2 + p_1}{p_{k+1}p_k p_{k-1} \cdots p_2 p_1}\right]$$

$$= \frac{1}{p_{k+1}p_k \cdots p_2 p_1}.$$

C. Hence it is true when $k = 2, 3, 4 \cdots$.

28. **Formulas for** (1^r). Formula {35} may now be written

$$(1^r) = \sum (-1)^{r-\rho} \frac{r!}{\pi_1! \cdots \pi_s!} \frac{1}{p_1 p_2 \cdots p_\rho} (p_1)(p_2) \cdots (p_\rho) \qquad \{37\}$$

or if the p's are ordered it may be written as

$$(1^r) = \sum (-1)^{r-\rho} \frac{r!}{\pi_1! \cdots \pi_s!} \frac{1}{p_1^{\pi_1} \cdots p_s^{\pi_s}} (p_1)^{\pi_1} \cdots (p_s)^{\pi_s} \qquad \{38\}$$

which is the formula previously given as {23} and {26}. In addition the check formulas {30} and {31} become

$$\sum (-1)^{r-\rho} \frac{r!}{p_1^{\pi_1} \cdots p_s^{\pi_s} \pi_1! \cdots \pi_s!} = 0 \qquad \{39\}$$

$$\sum \frac{r!}{p_1^{\pi_1} \cdots p_s^{\pi_s} \pi_1! \cdots \pi_s!} = r! \qquad \{40\}$$

These relations {39} and {40} correspond to statements of Cauchy (2), (I; 1; 252–3) and to later remarks of Cayley (D; 577). By dividing by $r!$, they become

$$\sum \frac{(-1)^{r-\rho}}{p_1^{\pi_1} \cdots p_s^{\pi_s} \pi_1! \cdots \pi_s!} = 0 \qquad \{41\}$$

$$\sum \frac{1}{p_1^{\pi_1} \cdots p_s^{\pi_s} \pi_1! \cdots \pi_s!} = 1 \qquad \{42\}.$$

The formula {38} is easily applied. Thus

$$(1^5) = \frac{5!}{5}(5) - \frac{5!}{4}(4)(1) - \frac{5!}{32}(3)(2) + \frac{5!}{32!}(3)(1)^2 + \frac{5}{222!}(2)^2(1)$$

$$- \frac{5!}{23!}(2)(1)^3 + \frac{5!}{5!}(1)^5$$

$$= 24(5) - 30(4)(1) - 20(3)(2) + 20(3)(1)^2 + 15(2)^2(1)$$

$$- 10(2)(1)^3 + (1)^5.$$

with

$$24 - 30 - 20 + 20 + 15 - 10 + 1 = 0$$

$$24 + 30 + 20 + 20 + 15 + 10 + 1 = 5!$$

We next write the formula (1^r) in such form that we use the principle of generalization from symmetry. If we multiply numerator and denominator of {38} by $(p_1-1)!^{\pi_1}(p_2-1)!^{\pi_2}\cdots(p_s-1)!^{\pi_s}$ we get

$$(1^r) = \sum (-1)^{r-\rho}(p_1-1)!^{\pi_1}(p_2-1)!^{\pi_2}\cdots(p_s-1)!^{\pi_s}\frac{r!\,(p_1)^{\pi_1}\cdots(p_s)^{\pi_s}}{(p_1!)^{\pi_1}(p_2!)^{\pi_2}\cdots(p_s!)^{\pi_s}\,\pi_1!\cdots\pi_s!}$$

which immediately becomes

$$(1^r) = \sum (-1)^{r-\rho}(p_1-1)!^{\pi_1}(p_2-1)!^{\pi_2}\cdots(p_s-1)!^{\pi_s}\binom{1^r}{p_1^{\pi_1}\cdots p_2^{\pi_2}}(p_1)^{\pi_1}\cdots(p_s)^{\pi_s} \quad \{43\}.$$

This somewhat formidable appearing formula is easy to apply. For example, in finding the value of (1^5) we write in one row all possible partitions of 5. In the next row we place the well known values of $\binom{1^r}{p_1^{\pi_1}\cdots p_s^{\pi_s}}$. In the next row we place the indicated products with proper signs. Thus

1^5	21^3	2^21	31^2	32	41	5
1	10	15	10	10	5	1
1	−1	+1	+2	−2	−6	+24

results in

$$(1^5) = (1)^5 - 10(2)(1)^3 + 15(2)^2(1) + 20(3)(1)^2 - 20(3)(2) - 30(4)(1) + 24(5)$$

as indicated above.

It is immediately recognized that formula {43} can be obtained from formula {3} by placing $P(1^r) = (1^r)$; $p_1^{\pi_1}\cdots p_s^{\pi_s}$ by $(p_1)^{\pi_1}\cdots(p_s)^{\pi_s}$ and $P_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$ by $(-1)^{r-\rho}(p_1-1)!^{\pi_1}\cdots(p_s-1)!^{\pi_s}$ and hence that formulas of Table I may be used in obtaining the values of (1^r).

29. Values of $(a_1\cdots a_r)$. The form of {43} also permits generalization from symmetry since the $\binom{1^r}{p_1^{\pi_1}\cdots p_s^{\pi_s}}$ equal values $(p_1)^{\pi_1}\cdots(p_s)^{\pi_s}$ are replaced by the $\binom{1^r}{p_1^{\pi_1}\cdots p_s^{\pi_s}}$ different values composing $T(p_1)^{\pi_1}\cdots(p_s)^{\pi_s}$ when the r units are replaced by the a's. It follows at once that

$$(a_1a_2\cdots a_r) = \sum(-1)^{r-\rho}(p_1-1)!^{\pi_1}\cdots(p_s-1)!^{\pi_s}T(p_1)^{\pi_1}\cdots(p_s)^{\pi_s} \quad \{44\}$$

where
$$r = p_1\pi_1 + p_2\pi_2 + \cdots + p_s\pi_s$$
and
$$\rho = \pi_1 + \pi_2 + \cdots + \pi_s.$$

As an illustration we write

$$(abc) = 2(a + b + c) - (a + b)(c) - (a + c)(b) - (b + c)(a) + (a)(b)(c)$$

as indicated earlier by {18} and

$$(a_1a_2a_3a_4) = -6T(4) + 2T(3)(1) + T(2)(2) - T(2)(1)^2 + T(1)^4$$

$$(a_1a_2a_3a_4a_5) = 24T(5) - 6T(4)(1) - 2T(3)(2) + 2T(3)(1)^2 + T(2)^2(1) - T(2)(1)^3 + T(1)^5$$

etc.

30. **Table of Values of $(a_1 \cdots a_r)$.** The values of the power products with $w \leqq 6$ are given in Table II which follows the general form of Table I. In fact Table II may be derived from Table I by placing every

$$P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} = (-1)^{r-\rho}(p_1 - 1)!^{\pi_1} \cdots (p_s - 1)!^{\pi_s}$$

as indicated in the next section.

31. **Use of Partition Formulas.** By comparing {44} with {4} we see that {44} can be obtained from {4} by placing

$$P(a_1a_2 \cdots a_3) = (a_1a_2 \cdots a_r)$$

$$P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} = (-1)^{r-\rho}(p_1 - 1)!^{\pi_1} \cdots (p_s - 1)!^{\pi_s}$$

and $$T_{p_1^{\pi_1} \cdots p_s^{\pi_s}} = T(p_1)^{\pi_1} \cdots (p_s)^{\pi_s}$$

It appears then that the values of any power product sum $(a_1a_2 \cdots a_r)$ can be obtained by writing the expansion of $P(a_1 \cdots a_r)$ and substituting as indicated. Thus since

$$P(321) = P_3 6 + P_{21} 51 + P_{21} 42 + P_{21} 33 + P_{111} 111$$

$$(321) = 2(6) - (5)(1) - (4)(2) - (3)(3) + (1)(1)(1).$$

It is also immediately apparent that Table II can be obtained from Table I by placing $P_{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ equal to $(-1)^{r-\rho}(p_1 - 1)!^{\pi_1} \cdots (p_s - 1)!^{\pi_s}$ and that the main results of Chapter I, including the recursion rule, are applicable to the present problem.

32. **Coefficients of Given Terms in the Expansion of Product Power Sums.** The methods of the last section are also useful in finding the coefficient of any term in the expansion. For example we wish to find the coefficient of (3)(2) in the expansion of (2111). We note that $P\binom{2111}{32} = P_{31} + 3P_{22}$ and that

TABLE II

Power product sums in terms of products of power sums when $W \leqq 6$

$W = 6$

	6	51	42	33	411	321	222	31^3	2^21^2	21^4	1^6
6	1										
51	−1	1									
42	−1		1								
33	−1			1							
411	2	−2	−1		1						
321	2	−1	−1	−1		1					
222	2		−3				1				
31^3	−6	6	3	2	−3	−3		1			
2^21^2	−6	4	5	2	−1	−4	−1		1		
21^4	24	−24	−18	−8	12	20	3	−4	−6	1	
1^6	−120	144	90	40	−90	−120	−15	40	45	−15	1

$W = 1$

	1
1	1

$W = 2$

	2	11
2	1	
11	−1	1

$W = 3$

	3	21	111
3	1		
21	−1	1	
111	2	−3	1

$W = 5$

	5	41	32	311	221	2111	11111
5	1						
41	−1	1					
32	−1		1				
311	2	−2	−1	1			
221	2	−1	−2		1		
2111	−6	6	5	−3	−3	1	
11111	24	−30	−20	20	+15	−10	1

$W = 4$

	4	31	22	211	1111
4	1				
31	−1	1			
22	−1		1		
211	2	−2	−1	1	
1111	−6	8	3	−6	1

the coefficient of (3)(2) is $P_{31} + 3P_{22}$ where $P_{31} = (-1)^{4-2}2! = 2$ and $P_{22} = (-1)^{4-2} = 1$. Hence the coefficient is $1 \cdot 2 + 3 \cdot 1 = 5$.

33. The Expansion of the Monomial Symmetric Function. If $a_1 \neq a_2 \neq a_3 \neq \cdots \neq a_r$ then $M(a_1 \cdots a_r) = (a_1 \cdots a_r)$ and previous results are applicable. If however the product power sum is of the form

$$(a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_h^{\alpha_h})$$

then

$$M(a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_h^{\alpha_h}) = \frac{1}{\alpha_1!\alpha_2! \cdots \alpha_h!} (a_1^{\alpha_1} \cdots a_h^{\alpha_h})$$

and

$$M(a_1^{\alpha_1} \cdots a_h^{\alpha_h})$$
$$= \frac{1}{\alpha_1!\alpha_2! \cdots \alpha_h!} \sum (-1)^{r-\rho}(p_1 - 1)!^{\pi_1} \cdots (p_s - 1)!^{\pi_s} T(p_1)^{\pi_1} \cdots (p_s)^{\pi_s} \quad \{45\}$$

For example

$$M(421) = 2(7) - (6)(1) - (5)(2) - (4)(3) + (4)(2)(1)$$
$$M(322) = (7) - (5)(2) - \tfrac{1}{2}(4)(3) + \tfrac{1}{2}(3)(2)(2).$$
$$M(2^2 1^2) = -\tfrac{3}{2}(6) + (5)(1) + \tfrac{5}{4}(4)(2) + \tfrac{1}{2}(3)(3) - \tfrac{1}{4}(4)(1)(1)$$
$$- (3)(2)(1) - \tfrac{1}{4}(2)(2)(2) + \tfrac{1}{4}(2)(2)(1)(1).$$

Study will show that the formula {45} is equivalent to one given by Faà de Bruno (C; 9) and later by Roe (7).

It is possible to use Table II in finding the expansion of the monomial symmetric functions. It is only necessary to multiply each term in the expansion of $(a_1 \cdots a_r)$ by $\dfrac{1}{\alpha_1! \cdots \alpha_h!}$.

The check formulas give, in the case of the monomial symmetric function: The sum of the coefficients in the expansion is 0.

The sum of the absolute values of the coefficients is $\dfrac{r!}{\alpha_1!\alpha_2! \cdots \alpha_h!}$.

The reader might compare the second of these checks with the results of Faà de Bruno (C; 14).

Tables giving the expansion of monomial symmetric function have been given. One by J. R. Roe (12; plate 18) includes all cases of weight $\leqq 10$.

34. **Previous Results.** Previous authors have studied the monomial symmetric function. Gordan has deduced a monomial symmetric function formula which is recommended by J. R. Roe (M; 24–33). MacMahon has given a general formula (K; II; 320) for expanding any monomial symmetric function in terms of power sums together with an operational method for its evaluation. O'Toole also has given a differential operator and showed how it could be applied in obtaining expansions (16; 115–130). O'Toole has also given a method of expanding symmetric functions in many variables by means of differential operators, (17).

Another method of attack was based upon the close relation existing between the elementary symmetric function and the determinant of the power sums. This has resulted in the expression of the monomial symmetric function in determinant form. Brioschi appears to have been the first (1854) to see how

a symbolic determinant could be used (3; 427) although he gave no proof. Bellavites tried in 1857, but obtained incorrect results (4). In 1876 Faà de Bruno made an attempt, but he too was in error (C; 10). In 1898 E. D. Roe, Jr. proved that Brioschi was right (7). Muir also gave a proof in 1908 (11; 5–9). The summation of determinants, rather than the symbolic determinant, was used by Hankel (6; 90–94) (L; III, 220).

The determinant of the power sums has been generalized in another way. A group of writers has studied the "immanents" of its matrix. D. E. Littlewood and A. R. Richardson have recently written a series of papers on this topic. One of these papers (18; 99–141) defined the term "immanents" and gave references to previous investigations dealing with this matrix.

It has been the aim of this chapter to present an easy development of the subject of the expansion of product power sums and monomial symmetric functions. This development is characterized by

1. The use of the formulas and tables of Chapter I in writing expansions of product power sums.
2. The use of product power sums in place of monomial symmetric functions which makes feasible
3. Generalization from symmetry.
4. References to previous work.

Chapter IV. The Double Expansion Theorem

In the present chapter we combine the multiplication expansion of Chapter II and the power product sum expansion of Chapter III into a new result which is to be known as the double expansion theorem. We show that this result may also be expressed in terms of the partition notation of Chapter I.

35. The Value of $K(a_1)(a_2)$. We know

$$(a_1)(a_2) = (a_1 + a_2) + (a_1 \cdot a_2)$$

and if we multiply $(a_1 + a_2)$ by k_2 and $(a_1 \cdot a_2)$ by k_{11} we have a new expression which we designate by $K(a_1)(a_2)$.

$$K(a_1)(a_2) = k_2(a_1 + a_2) + k_{11}(a_1 \cdot a_2) \qquad \{46\}$$

Since

$$(a_1a_2) = (a_1)(a_2) - (a_1 + a_2)$$

$$K(a_1)(a_2) = k_2(a_1 + a_2) + k_{11}[(a_1)(a_2) - (a_1 + a_2)]$$

$$K(a_1)(a_2) = (k_2 - k_{11})(a_1 + a_2) + k_{11}(a_1)(a_2)$$

which can be written

$$K(a_1)(a_2) = K_2(a_1 + a_2) + K_{11}(a_1)(a_2) \qquad \{47\}$$

if $k_2 - k_{11} = K_2$ and $K_{11} = k_{11}$

36. The Value of $K(a_1)(a_2)(a_3)$. We know from {12} that

$$(a_1)(a_2)(a_3) = T(3) + T(21) + T(111)$$

and we define $K(a_1)(a_2)(a_3) = k_3T(3) + k_{21}T(21) + k_{111}T(111)$. Inserting the values $T_3 = (a_1 + a_2 + a_3)$, $T_{21} = (\overline{a_1 + a_2} \cdot a_3) + (\overline{a_1 + a_3} \cdot a_2) + (\overline{a_2 + a_3} \cdot a_1)$, $T_{111} = (a_1a_2a_3)$ and reducing to power sums by {44}, we get

$$K(a_1)(a_2)(a_3) = (k_3 - 3k_{21} + 2k_{111})(a_1 + a_2 + a_3) + (k_{21} - k_{111})$$
$$\{(a_1 + a_2)(a_3) + (a_1 + a_3)(a_2) + (a_2 + a_3)(a_1)\} + k_{111}(a_1)(a_2)(a_3)$$

which may be written

$$K(a_1)(a_2)(a_3) = K_3(a_1 + a_2 + a_3) + K_{21}\{(a_1 + a_2)(a_3)$$
$$+ (a_1 + a_3)(a_2) + (a_2 + a_3)(a_1)\} + K_{111}(a_1)(a_2)(a_3) \qquad \{48\}$$

where $K_3 = k_3 - 3k_{21} + 2k_{111}$, $K_{21} = k_{21} - k_{111}$, $K_{111} = k_{111}$.

37. Definition of $K(a_1)(a_2) \cdots (a_r)$. We define

$$K(a_1)(a_2) \cdots (a_r) = \sum k_{p_1^{\pi_1} \cdots p_s^{\pi_s}} T(p_1^{\pi_1} \cdots p_s^{\pi_s}) \qquad \{49\}$$

where $T(p_1^{\pi_1} \cdots p_s^{\pi_s})$ is composed of $\binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ power product sums. We wish to find the value $K(a_1)(a_2) \cdots (a_r)$ in terms of power sums. This involves the expansion of each power product sum in terms of power sums and then the collection of the results. This algebraic process is to be called the double expansion process and the theorem which results, the double expansion theorem.

38. Special Cases of the Theorem. The results {47} and {48} are special cases of the double expansion theorem when $r = 2, 3$. When $r = 1$ it is evident that $K(a_1) = K_1(a_1) = k_1(a_1)$. {50}

The results {50}, {48}, and {49} may be written symbolically by

$$K(a_1) = K_1T(1)$$
$$K(a_1)(a_2) = K_2T(2) + K_{11}T(1)^2 \qquad \{51\}$$
$$K(a_1)(a_2)(a_3) = K_3T(3) + K_{21}T(2)(1) + K_{111}T(1)^3$$

It can also be shown, with a much more extensive use of the results of Chapters II and III, that

$$K(a_1)(a_2)(a_3)(a_4) = K_4T(4) + K_{31}T(3)(1) + K_{22}T(2)^2$$
$$+ K_{211}T(2)(1)^2 + K_{1111}T(1)^4 \qquad \{52\}$$

$$K(a_1) \cdots (a_5) = K_5T(5) + K_{41}T(4)(1) + K_{32}T(3)(2)$$
$$+ K_{311}T(3)(1)^2 + K_{221}T(2)^2(1) + K_{2111}T(2)(1)^3$$
$$+ K_{11111}T(1)^5 \qquad \{53\}$$

where

$$\left.\begin{aligned}
K_4 &= k_4 - 4k_{31} - 3k_{22} + 12k_{211} - 6k_{1111}\\
K_{31} &= k_{31} - 3k_{211} + 2k_{1111}\\
K_{22} &= k_{22} - 2k_{211} + k_{1111}\\
K_{211} &= k_{211} - k_{1111}\\
K_{1111} &= k_{1111}
\end{aligned}\right\} \quad \{54\}$$

and

$$\left.\begin{aligned}
K_5 &= k_5 - 5k_{41} - 10k_{32} + 20k_{311} + 30k_{221} - 60k_{2111} + 24k_{11111}\\
K_{41} &= k_{41} - 4k_{311} - 3k_{221} + 12k_{2111} - 6k_{11111}\\
K_{32} &= k_{32} - k_{311} - 3k_{221} + 5k_{2111} - 2k_{11111}\\
K_{311} &= k_{311} - 3k_{2111} + 2k_{11111}\\
K_{221} &= k_{221} - 2k_{2111} + k_{11111}\\
K_{2111} &= k_{2111} - k_{11111}\\
K_{11111} &= k_{11111}
\end{aligned}\right\} \quad \{55\}$$

We may say then that, for $r < 6$

$$K(a_1) \cdots (a_r) = \sum k_{p_1^{\pi_1}\cdots p_s^{\pi_s}}\, T(p_1^{\pi_1} \cdots p_s^{\pi_s})$$
$$= \sum k_{p_1^{\pi_1}\cdots p_s^{\pi_s}}\, T(p_1)^{\pi_1} \cdots (p_s)^{\pi_s} \quad \{56\}$$

where $K_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$ is defined by the relations {47}, {48}, {54}, and {55}. In examining the value of K_r we note

$$\begin{aligned}
K_1 &= k_1\\
K_2 &= k_2 - k_{11}\\
K_3 &= k_3 - 3k_{21} + 2k_{111}\\
K_4 &= k_4 - 4k_{31} - 3k_{22} + 12k_{211} - 6k_{1111}\\
K_5 &= k_5 - 5k_{41} - 10k_{32} + 20k_{311} + 30k_{221} - 60k_{2111} + 24k_{11111}
\end{aligned}$$

and that these are given, for $r < 6$ by

$$K_r = \sum (-1)^\rho(\rho - 1)! \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} k_{p_1^{\pi_1}\cdots p_s^{\pi_s}} \quad \{57\}$$

It is further to be noted that {57} can be obtained from {3} by placing $P(1^r) = K_r$, $p_1^{\pi_1} \cdots p_s^{\pi_s}$ by $k_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$, and $P_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$ by $(-1)^\rho(\rho - 1)!$ Hence the last rows of Table I may be used in writing the values of K_r. Thus from

$$P(1^3) = P_3(3) + 3P_{21}(21) + P_{111}(1)^3$$

we get

$$K_3 = k_3 - 3k_{21} + 2k_{111}\,.$$

It is further evident that if $\overline{K_3K_2} = \overline{(k_3 - 3k_{21} + 2k_{111})(k_2 - k_{11})}$ indicates multiplication by suffixing of subscripts that $\overline{K_3K_2} = k_{32} - k_{311} - 3k_{221} + 5k_{2111} - 2k_{11111} = K_{32}$

and in general it can be shown that for $r < 6$

$$\begin{aligned} K_{r_1r_2} &= \overline{K_{r_1}K_{r_2}} \\ K_{r_1r_2r_3} &= \overline{K_{r_1}K_{r_2}K_{r_3}} \\ &\text{etc.} \end{aligned} \tag{58}$$

so that all values $K_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$ may be obtained by symbolic multiplication of equations {57}.

The method of this section can be used in demonstrating that the results {56}, {57}, and {58} hold also when $r = 6, 7, 8 \cdots$, but the amount of algebraic manipulation increases enormously with each increase in r. We establish these results, for all integral values of r, by a more general approach.

39. A More General Definition. We provide a more general definition of $K(a_1) \cdots (a_r)$ by letting the subscripts of the k's agree with parts of the given partition rather than with its complete order. Thus

$$K'(a_1)(a_2) = k_{a_1+a_2}(a_1 + a_2) + k_{a_1a_2}(a_1a_2)$$

and in general, if $q_1^{x_1} \cdots q_t^{x_t}$ represents any ρ part partition having complete order $p_1^{\pi_1} \cdots p_s^{\pi_s}$, then we may define

$$K'(a_1)(a_2) \cdots (a_r) = \sum k_{q_1^{x_1}\cdots q_t^{x_t}}\,(q_1^{x_1} \cdots q_t^{x_t}) \tag{59}$$

where the summation holds, not only for every different complete order as does {49}, but for every possible partition. By {44} $(q_1^{x_1} \cdots q_t^{x_t})$ may be written as

$$(q_1^{x_1} \cdots q_t^{x_t}) = \sum(-1)^{\rho-\sigma}(d_1 - 1)!(d_2 - 1)! \cdots (d_\sigma - 1)!T(d_1) \cdots (d_\sigma)$$

where $d_1 + d_2 + \cdots + d_\sigma = \rho$ and where groups of the d's may be alike. If $(w_1)(w_2) \cdots (w_\sigma)$ is one of the products of power sums having the complete order $(d_1 \cdots d_\sigma)$ we may write

$$(q_1^{x_1} \cdots q_t^{x_t}) = \sum(-1)^{\rho-\sigma}(d_1 - 1)! \cdots (d_\sigma - 1)!(w_1)(w_2) \cdots (w_\sigma) \tag{60}$$

where

$$q_1x_1 + \cdots + q_tx_t = w = w_1 + w_2 + \cdots + w_\sigma$$

and

$$x_1 + \cdots + x_t = \rho$$

and where the summation sign holds not only for every complete order $d_1 \cdots d_g$, but for all power sum partition products $(w_1)(w_2) \cdots (w_g)$.

The insertion of {60} in {59} gives

$$K'(a_1)(a_2) \cdots (a_r) = \sum k_{q_1^{\pi_1} \cdots q_t^{\pi_t}} \sum (-1)^{\rho-g}(d_1-1)! \cdots (d_g-1)!(w_1) \cdots (w_g) \quad \{61\}$$

40. **Value of K'_w.** The notation of K'_w is used to indicate the coefficient of the power sum $(w) = (a_1 + a_2 + \cdots + a_r)$ in the expansion of {61}. In this case $d_1 = \rho$ and $g = 1$ so that

$$K'_w = \sum k_{q_1^{\pi_1} \cdots q_t^{\pi_t}} (-1)^{\rho-1}(\rho-1)! \quad \{62\}$$

which may be written more symbolically as

$$K'_w = \sum (-1)^{\rho-1}(\rho-1)!\, k_{\pi_w} \quad \{63\}$$

where π_w represents any algebraic partition of $a_1 + \cdots + a_r$ and ρ indicates the number of its parts.

41. **Products of K''s.** The notation $\overline{K'_{w_1} K'_{w_2}}$ is used to indicate the product of K'_{w_1} by K'_{w_2} if the rule of multiplication is the suffixing of the subscripts of the k's in the expansion of K'_{w_1} and K'_{w_2}. Thus

$$\begin{aligned} \overline{K'_{a_1+a_2} \cdot K'_{a_3}} &= \overline{(k_{a_1+a_2} - k_{a_1 a_2})(k_{a_3})} \\ &= k_{a_1+a_2 \cdot a_3} - k_{a_1 a_2 a_3} \end{aligned}$$

More generally, if we write, from {63}

$$\begin{aligned} K'_{w_1} &= \sum (-1)^{d_1-1}(d_1-1)!\, k_{\pi_{w_1}} \\ K'_{w_2} &= \sum (-1)^{d_2-g}(d_2-1)!\, k_{\pi_{w_2}} \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ K'_{w_g} &= \sum (-1)^{d_g-1}(d_g-1)!\, k_{\pi_{w_g}} \end{aligned}$$

and use multiplication by suffixing of subscripts we have

$$\overline{K'_{w_1} K'_{w_2} \cdots K'_{w_g}} = \sum (-1)^{\rho-g}(d_1-1)! \cdots (d_g-1)\, k_{\pi_{w_1} \cdots \pi_{w_g}} \quad \{64\}$$

where $\rho = d_1 + d_2 + \cdots + d_g$ and the summation holds for every partition which can be formed by combining any algebraic partition of w_1, any partition of $w_2, \cdots$, any partition of w_g.

42. **The Coefficient of $(w_1)(w_2) \cdots (w_g)$.** The coefficients of any specific product of power sums $(w_1)(w_2) \cdots (w_g)$ is from {61}

$$K'_{w_1 w_2 \cdots w_g} = \sum k_{q_1^{\pi_1} \cdots q_t^{\pi_t}} (-1)^{\rho-g}(d_1-1)! \cdots (d_g-1)! \quad \{65\}$$

where the summation holds, not only for the partitions of $a_1 + a_2 + \cdots + a_r$, but for the partitions $\pi_{w_1}, \pi_{w_2}, \cdots, \pi_{w_g}$ since these partitions can be combined to form $(w_1)(w_2) \cdots (w_g)$. Hence {65} becomes

$$K'_{w_1w_2\cdots w_g} = \sum (-1)^{\rho-g}(d_1 - 1)! \cdots (d_g - 1)!\, k_{\pi_{w_1}\cdots\pi_{w_g}} \qquad \{66\}$$

and it is immediately seen that the right hand expressions of {66} and {64} are the same and hence that

$$K'_{w_1w_2\cdots w_g} = \overline{K'_{w_1}K'_{w_2}\cdots K'_{w_g}}$$

as expected from (58).

We can now say that

$$\begin{aligned} K'(a_1)(a_2)\cdots(a_r) &= \sum k_{q_1^{\pi_1}\cdots q_t^{\pi_t}}\,(q_1^{\pi_1}\cdots q_t^{\pi_t}) \\ &= \sum K'_{w_1w_2\cdots w_g}(w_1)(w_2)\cdots(w_g) \qquad \{67\} \end{aligned}$$

where

$$K'_w = \sum (-1)^{\rho-1}(\rho - 1)!\, k_{\pi_w} \qquad \{68\}$$

and

$$K'_{w_1w_2\cdots w_g} = \overline{K'_{w_1}K'_{w_2}\cdots K'_{w_g}} \qquad \{69\}$$

Relations {67}, {68} and {69} constitute the general double expansion theorem.

43. The Double Expansion Theorem. The case of the double expansion theorem in which we are especially interested is that in which the coefficients of all similar power sum products are the same, i.e., $k_{q_1^{\pi_1}\cdots q_s^{\pi_s}}$ is a function of the complete order indicated by $k_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$. In this case {68} becomes

$$K_w = \sum (-1)^{\rho}(\rho - 1)! \binom{1^r}{p_1^{\pi_1}\cdots p_s^{\pi_s}} k_{p_1^{\pi_1}\cdots p_s^{\pi_s}} \qquad \{70\}$$

where the summation holds for all possible complete orders. Suppose now that the r algebraic expressions, $a_1, a_2, \cdots, a_r$ are all unity then {69} becomes

$$K_r = \sum (-1)^{\rho-1}(\rho - 1)! \binom{1^r}{p_1^{\pi_1}\cdots p_s^{\pi_s}} k_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$$

and we find that $K_w = K_r$. We may then write {67}, {68} and {69} as

$$K(a_1)\cdots(a_r) = \sum k_{p_1^{\pi_1}\cdots p_s^{\pi_s}}\,(p_1^{\pi_1}\cdots p_s^{\pi_s}) = \sum K_{r_1\cdots r_g}\, T(r_1)\cdots(r_g) \qquad \{71\}$$

where

$$K_r = \sum (-1)^{\rho-1}(\rho - 1)! \binom{1^r}{p_1^{\pi_1}\cdots p_s^{\pi_s}} k_{p_1^{\pi_1}\cdots p_s^{\pi_s}} \qquad \{72\}$$

and

$$K_{r_1r_2\cdots r_g} = \overline{K_{r_1}K_{r_2}\cdots K_{r_g}} \qquad \{73\}$$

Now $r_1r_2 \cdots r_q$ indicates any grouping of the a's, and hence any complete order of $a_1 + a_2 + \cdots + a_r$. So {71} may be written, with a slight change of notation as

$$K(a_1) \cdots (a_r) = \sum k_{p_1^{\pi_1} \cdots p_s^{\pi_s}} T(p_1^{\pi_1} \cdots p_s^{\pi_s}) = \sum K_{p_1^{\pi_1} \cdots p_s^{\pi_s}} T(p_1)^{\pi_1} \cdots (p_s)^{\pi_s} \quad \{74\}$$

The relations {74}, {72} and {73} are the desired generalizations of {56}, {57} and {58} and hold for all positive integral values of r.

The double expansion theorem provides a method of writing out the result of the double expansion process without going through the work involved in the process. Thus

$$\begin{aligned} K(3)(2)(1) &= K_3(6) + K_{21}\{(5)(1) + (4)(2) + (3)(3)\} \\ &+ K_{111}(3)(2)(1) = (k_3 - 3k_{21} + 2k_{111})(6) \\ &+ (k_{21} - k_{111})\{(5)(1) + (4)(2) + (3)(3)\} + k_{111}(3)(2)(1) \end{aligned} \quad \{75\}$$

44. The Double Expansion Theorem and Partition Notation. It is immediately evident that {74} can be obtained from {4} if $P(a_1 \cdots a_r)$ is replaced by $K(a_1) \cdots (a_r)$, if $P_{p_1^{\pi_1} \cdots p_s^{\pi_s}}$ is replaced by $K_{p_1^{\pi_1} \cdots p_s^{\pi_s}}$, and if $p_1^{\pi_1} \cdots p_s^{\pi_s}$ is replaced by $T(p_1)^{\pi_1} \cdots (p_s)^{\pi_s}$. It follows at once that the entire theory of Chapter I,—table, recursion formula, etc.,—is applicable to double expansion theory. For example {75} above is obtained from

$$P(321) = P_3 6 + P_{21}\{51 + 42 + 33\} + P_{111}321$$

simply by replacing the K's by the P's and enclosing the parts in parentheses. We can as well use P's as K's to represent the double expansion theorem and hence have available a list of double expansion formulas when $w \leqq 6$. We also have available a recursion property for writing double expansions beyond the scope of the table. Thus for example, the illustration at the end of section 9 may be interpreted as a statement of the double expansion theorem when $a_1 = 3$, $a_2 = 2$, $a_3 = 2$, $a_4 = 1$.

45. The Case of Equal Powers. In case $a_1 = a_2 = a_3 = \cdots = a$, {74} reduces to {3'} of Chapter I with

$$P_r = \sum (-1)^{\rho-1}(\rho - 1)! \binom{1'}{p_1^{\pi_1} \cdots p_s^{\pi_s}} k_{p_1^{\pi_1} \cdots p_s^{\pi_s}} \quad \{76\}$$

and

$$P_{r_1 r_2 \cdots r_q} = \overline{P_{r_1} P_{r_2} \cdots P_{r_q}}.$$

Formula {74} also reduces to {3} when $a_1 = a_2 = \cdots = a_r = 1$.

46. Special Values of $K_{p_1^{\pi_1} \cdots p_s^{\pi_s}}$.

A. $k_{p_1^{\pi_1} \cdots p_s^{\pi_s}} = 1$. In this case the coefficients are all unity and

$$P(a_1)(a_2) \cdots (a_r) = (a_1)(a_2) \cdots (a_r)$$

It follows that $P_r = 0$ and that $P_{p_1^{\pi_1}\cdots p_s^{\pi_s}} = 0$ except that $P_r = 1$. Placing $P_r = 0$ and $k_{p_1^{\pi_1}\cdots p_s^{\pi_s}} = 1$ in {72} or its equivalent {76} we have, when $r > 1$

$$0 = \sum (-1)^{\rho-1}(\rho-1)!\begin{pmatrix} 1^r \\ p_1^{\pi_1} \cdots p_s^{\pi_s}\end{pmatrix} \qquad \{77\}$$

where the summation holds for every partition of r. This formula should be compared with {39} and {40}. When $r = 4$ and the partitions are

	4,	31,	22,	211,	1^4		
{77} gives	1	-4	-3	$+12$	-6	=	0
{39} gives	-6	$+8$	$+3$	-6	$+1$	=	0
{40} gives	6	$+8$	$+3$	$+6$	$+1$	=	4!

The equivalent of {77} was first given by Cayley (D; 576) who at the same time noted the similarity to {39}.

It follows immediately that the sum of the coefficients in the expansion of $P_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$, except P_{1^r}, is 0, for the sum of the coefficients of $P_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$ is the sum of the coefficients of $\overline{(P_{p_1})^{\pi_1} \cdots (P_s)^{\pi_s}}$ and is 0. For example the sum of the coefficient of $P_{32} = k_{32} - k_{311} - 3k_{221} + 5k_{2111} - 2k_{11111}$ is 0.

Since the coefficients of $(\mu_1')^{\pi_1} \cdots (\mu_s')^{\pi_s}$ (19; 25) in the expansion of Thiele half invariants are $(-1)^{\rho-1}(\rho-1)!\begin{pmatrix} 1^r \\ p_1^{\pi_1} \cdots p_s^{\pi_s}\end{pmatrix}$ it follows from {77} that the sum of these coefficients is 0.

B. $k_{p_1^{\pi_1}\cdots p_s^{\pi_s}} = \dfrac{n^{(\rho)}}{N^{(\rho)}}$. In this case all terms having the same number of parts, ρ, have the same coefficients. If we indicate $\dfrac{n^{(\rho)}}{N^{(\rho)}}$ by $\rho_1, \rho_2, \cdots,$ when $\rho = 1, 2, \cdots,$ {57}, {76} become

$$\begin{aligned}
P_1 &= \rho_1 \\
P_2 &= \rho_1 - \rho_2 \\
P_3 &= \rho_1 - 3\rho_2 + 2\rho_3 \\
P_4 &= \rho_1 - 7\rho_2 + 12\rho_3 - 6\rho_4 \\
P_5 &= \rho_1 - 15\rho_2 + 50\rho_3 - 60\rho_4 + 24\rho_5 . \\
&\text{etc.}
\end{aligned}$$

which are the formulas which have been used by Carver (15) and O'Toole (16).

Many other additional cases can be obtained by giving different values to $k_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$, but a discussion of these is hardly justified here as the case in which $k_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$ is a function of the number of parts, ρ, is to be used in Part II.

47. Relation to Previous Results. No general statement of the double expansion theorem has previously been given although the special case $K(a^r)$

has been developed by Carver (15) and O'Toole (16). Their results are further restricted to the special case (B) of section 46. The application of the double expansion theorem in this case is very useful in studying sampling from a finite universe as Carver has shown and as is demonstrated in Part II.

Most writers who have worked on the problem of moments of moments have gone through the double expansion process, but Carver was the first to note that the result of the process can be written in terms of the P polynomials above. It seems appropiate therefore to refer to these P polynomials of the coefficients as Carver polynomials.

Chapter V. The Multipartition and Multivariate Formulas

It is the purpose of this chapter to show how the results of Chapters I, II' III, and IV may be extended to the case of different variables.

48. **Multipartitions. Tables.** Formula {4} is still applicable if we let the a_1 units be the units of one quantity, the a_2 units to be the units of a second quantity, etc. Thus for example the formula $P(a_1a_2a_3)$ may be used to represent the precise number of ways in which a_1 apples, a_2 pears and a_3 peaches can be formed into groups without breaking up the groups of apples, pears, and peaches.

Various conventions for representing multipartitions of this type have been used. We adopt the one in which the individual partitions are written in successive columns. The partitions of the first number are combined with the partitions of the second number to form all possible multipartitions. Thus the multipartite number 111 has the partitions

111	110	101	011	100
	001	010	100	010
				001

where the parts are given in the rows. It is desired to show the number of ways in which any one of those partitions may be combined to form partitions of fewer parts. Thus

$$P\begin{pmatrix}100\\010\\001\end{pmatrix} = P_3 111 + P_{21}\begin{matrix}110\\001\end{matrix} + P_{21}\begin{matrix}101\\010\end{matrix} + P_{21}\begin{matrix}011\\100\end{matrix} + P_{111}\begin{matrix}100\\010\\001\end{matrix}$$

This is obtained from $P(a_1a_2a_3)$ by placing $a_1 = 1_1$, $a_2 = 1_2$, $a_3 = 1_3$, and could be written from {4} as

$$P(1_11_21_3) = P_3(1_1 + 1_2 + 1_3) + P_{21}\{\overline{1_1 + 1_2}\cdot 1_3 + \overline{1_1 + 1_3}\cdot 1_2 + \overline{1_2 + 1_3}\cdot 1_1\} + P_{111} 1_1\cdot 1_2\cdot 1_3.$$

Similarly

$$P\begin{pmatrix}10\\10\\01\\01\end{pmatrix} = P_4\begin{matrix}2\\2\end{matrix} + 2P_{31}\begin{matrix}21\\01\end{matrix} + 2P_{31}\begin{matrix}12\\10\end{matrix} + P_{22}\begin{matrix}20\\02\end{matrix} + 2P_{11}\begin{matrix}11\\11\end{matrix} + P_{211}\begin{matrix}20\\01\\01\end{matrix} + P_{211}\begin{matrix}02\\10\\10\end{matrix}$$

$$+ 4P_{211}\begin{matrix}11\\10\\01\end{matrix} + P_{1111}\begin{matrix}10\\10\\01\\01\end{matrix}$$

is a special case of $P(a_1a_2a_3a_4)$ where $a_1 = 1_1$, $a_2 = 1_1$, $a_3 = 1_2$, $a_4 = 1_2$. Formula {4} is also true where the a_i units are not of the same kind. Thus

$$P(a_1a_2) = P_2(a_1 + a_2) + P_{11}(a_1a_2)$$

gives

$$P\begin{pmatrix}11\\10\end{pmatrix} = P_2 21 + P_{11}\begin{matrix}11\\10\end{matrix} \qquad \text{when} \quad a_1 = 1_1 + 1_2 \quad \text{and} \quad a_2 = 1_1.$$

TABLE III

The Multipartite Number 11

	11	10 01
11	P_1	
10 01	P_2	P_{11}

The Multipartite Number 111

	111	110 001	101 010	011 100	100 010 001
111	P_1				
110 001	P_2	P_{11}			
101 010	P_2		P_{11}		
011 100	P_2			P_{11}	
100 010 001	P_3	P_{21}	P_{21}	P_{21}	P_{111}

TABLE III—*Continued*

The Multipartite Number 22

	22	21 01	12 10	20 02	11 11	20 01 01	02 10 10	11 10 01	10 10 01 01
22	P_1								
21 01	P_2	P_{11}							
12 10	P_2		P_{11}						
20 02	P_2			P_{11}					
11 11	P_2				P_{11}				
20 01 01	P_3	$2P_{21}$		P_{21}		P_{111}			
02 10 10	P_3		$2P_{21}$	P_{21}			P_{111}		
11 10 01	P_3	P_{21}	P_{21}		P_{21}			P_{111}	
10 10 01 01	P_4	$2P_{31}$	$2P_{31}$	P_{22}	$2P_{22}$	P_{211}	P_{211}	$4P_{211}$	P_{1111}

Tables can be made for the partitions of the various multipartite numbers. In Table III are presented values for the numbers 11, 111, 22.

When the units are indistinguishable 11 condenses to the $w = 2$ part of Table I.

When the units are indistinguishable 111 condenses to the $w = 3$ part of Table I.

When the units are alike 22 condenses to the $w = 4$ part of Table I.

49. Multivariate Distributions. The chief results of Chapters II, III, IV also hold for multivariate distributions. Some additional definitions are neces-

sary. We suppose that the N variates $x_1, x_2, \cdots, x_N$ are replaced by the Nr variates of the array

$$\begin{array}{l} {}_1x_1,\ {}_1x_2,\ \cdots,\ {}_1x_N \\ {}_2x_1,\ {}_2x_2,\ \cdots,\ {}_2x_N \\ \cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \\ {}_rx_1,\ {}_rx_2,\ \cdots,\ {}_rx_N \end{array} \tag{78}$$

where the presubscript represents the variable. The power sums become

$$(a_1) = {}_1x_1^{a_1} + {}_1x_2^{a_1} + \cdots + x_N^{a_1} = \sum {}_1x_i^{a_1}$$
$$(a_2) = {}_2x_1^{a_2} + {}_2x_2^{a_2} + \cdots + {}_2x_N^{a_2} = \sum {}_2x_i^{a_2}$$

It is not necessary to utilize the presubscript since it is precisely the subscript of the a. That is the power sum (a_k) is defined by $\sum x_i^{a_k}$. Similarly $(a_1 a_2) = \sum_{i \neq j} {}_1x_i^{a_1}\, {}_2x_i^{a_2}$ can be written as $(a_1a_2) = \sum_{i \neq j} x_i^{a_1}\, x_j^{a_2}$ without introducing ambiguity. In general {6} as well as {4}, now holds for the multivariate case. It follows at once that the results of Chapters II, III, IV can be written for the multivariate case by means of the formulas of Chapter I as indicated by the previous section. Thus the formula for $P(1_11_11_21_2)$ may be written as [Table III]

$$\begin{aligned} P[\overline{10}\cdot\overline{10}\cdot\overline{01}\cdot\overline{01}] &= P_4\overline{22} + 2P_{31}\overline{21}\cdot\overline{01} + 2P_{21}\overline{12}\cdot\overline{10} + P_{22}\overline{20}\,\overline{02} + 2P\,\overline{11}\,\overline{11} \\ &\quad + P_{211}\overline{20}\,\overline{01}\,\overline{01} + P_{211}\overline{02}\,\overline{10}\,\overline{10} + 4P_{211}\overline{11}\,\overline{10}\,\overline{01} + P_{1111}\overline{10}\,\overline{10}\,\overline{01}\,\overline{01} \end{aligned}$$

and can be interpreted as:

$$\begin{aligned} (10)^2(01)^2 &= (\overline{22}) + 2(\overline{21}\cdot\overline{01}) + 2(\overline{12}\cdot\overline{10}) + (\overline{20}\cdot\overline{02}) + 2(\overline{11}\cdot\overline{11}) + (\overline{20}\cdot\overline{01}\cdot\overline{01}) \\ &\quad + (\overline{02}\cdot\overline{10}\cdot\overline{10}) + (\overline{11}\cdot\overline{10}\cdot\overline{01}) + (\overline{10}\cdot\overline{10}\cdot\overline{01}\cdot\overline{01}) \end{aligned}$$

by {12} of Chapter II. It may also be interpreted as

$$\begin{aligned} (\overline{10}\cdot\overline{10}\cdot\overline{01}\cdot\overline{01}) &= -6(22) + 4(21)(01) + 4(12)(10) + (20)(02) \\ &\quad + 2(11)(11) - (20)(01)(01) - (02)(10)(10) \\ &\quad - 4(11)(10)(01) + (10)(10)(01)(01) \end{aligned}$$

by {44} of Chapter II. It can also be interpreted as a double expansion by means of section 44 where the values of the P's are given by the usual

$$P_r = \sum (-1)^{\rho-1}(\rho - 1)!\begin{pmatrix} & 1^r & \\ p_1^{\pi_1} & \cdots & p_s^{\pi_s} \end{pmatrix} k_{p_1^{\pi_1}\cdots p_s^{\pi_s}}$$

$$P_{r_1\cdots r_\rho} = \overline{P_{r_1}P_{r_2}\cdots P_{r_\rho}}$$

50. **Summary.** It is apparent that {4} not only expresses (a) the number of ways in which the parts of one partition may be collected to form the parts

of another partition, (b) the formula for expanding products of power sums in terms of power product sums, (c) the formula expanding power product sums in terms of power sums, and (d) the formula for double expansions, but also that it can be used to make similar expansions in the case of multivariate distributions.

BIBLIOGRAPHY

(1) BINET: "Mémoire sur un système de formules analytiques etc." Journ. de l'Ec. Polyt., IX (1812), cah 16, pp. 280–302.

(2) CAUCHY: "Note sur la formation des fonctions alternées qui servent à résoudre le problème de l'élimination." Comptes Rendus, 12 (1841), pp. 414–426.

(3) BRIOSCHI: "Sulle funzioni simmetriche delle radici di una equazione." Annali di Tortolini, 5 (1854), pp. 422–428.

(4) BELLAVITIS: "Sposizione elementare della teorica dei determinante." Memorie .... Istituto Veneto, 7 (1857), pp. 67–144.

(5) MOLA: "Soluzione della quistione 5, 6, 7." Giornale di Matematiche, 3 (1865), pp. 190–201.

(6) HANKEL: "Darstellung symmetrischer Functionen durch die Potenzsummen." Crelle's Journ., 67 (1865), pp. 90–94.

(7) E. D. ROE, JR.: "Note on a formula of symmetric functions." American Mathematical Monthly, 5 (1898), pp. 161–164.

(8) E. D. ROE, JR.: "On the transcendental form of the resultant." American Mathematical Monthly, 7 (1900), pp. 59–66.

(9) MUIRHEAD: "Some proofs of Newton's theorems on the sums of powers of roots." Edinburgh Mathematical Society Proceedings, 23 (1905), pp. 66–70.

(10) MUIRHEAD: "A proof of Waring's expression for $\Sigma\alpha^r$ in terms of the coefficients of the equation." Edinburgh Mathematical Society Proceedings, 23 (1905), pp. 71–74.

(11) MUIR: "Waring's expression for symmetric functions in terms of sums of like powers." Proceedings Edinburgh Math. Society, 27 (1909), pp. 5–9.

(12) A. A. TCHOUPROFF: "On the mathematical expectation of moments of frequency distributions." Biom., 12 (1918), pp. 140–169.

(13) A. E. R. CHURCH: "On the moments of the distribution of the squared standard deviations, etc." Biom., 17 (1925), pp. 79–83.

(14) A. E. R. CHURCH: "On the means and squared standard deviations of small samples." Biom., 18 (1926), pp. 321–394.

(15) H. C. CARVER: "The fundamentals of sampling." Annals of Mathematical Statistics, 1 (1930), pp. 101–121.

(16) A. L. O'TOOLE: "On symmetric functions and symmetric functions of symmetric functions." Annals of Mathematical Statistics, 2 (1931), 101–149.

(17) A. L. O'TOOLE: "On symmetric functions of more than one variable and of frequency functions." Annals of Mathematical Statistics, 3 (1932), pp. 56–63.

(18) D. E. LITTLEWOOD AND A. R. RICHARDSON: "Group characters and algebra." Phil. Trans. Roy. Soc., A 233 (1934), pp. 99–141.

(19) P. S. DWYER: "Moments of any rational integral isobaric sample moment function." Annals of Mathematical Statistics, vol. viii, no. 1, Mar. 1937, pp. 21–65.

A. PAOLI: "Elementi di algebra." Supplemento Opuscolo II, 1804.

B. HIRSCH: "Examples, formulae, and calculations on the literal calculus and algebra." Translated from the German by J. A. Ross (1827).

C. FAÀ DE BRUNO: "Théorie des formes binaires" (1876).

D. CAYLEY: "Collected Mathematical Papers," VII (1894).

E. BURNSIDE AND PANTON: "Theory of equations" (1881). Reference is to the seventh edition (1912).
F. CHRYSTAL: "Algebra" (1886). Reference is to the fifth edition.
G. WHITWORTH: "Choice and chance," 4th edition (1886).
H. T. N. THIELE: "Almindelig Iagttagelseslaere" (1889).
I. THOMAS MUIR: "Theory of determinants."
J. M. BOCHER: "Introduction to higher algebra" (1907).
K. MACMAHON: "Combinatory analysis" (1915–16).
L. THOMAS MUIR: "Contributions to the history of determinants" (1900-1920).
M. JOSEPHINE ROE: "*Interfunctional expressibility problems of symmetric functions*" (1931).
N. JOSEPHINE ROE: "Interfunctional expressibility tables of symmetric functions." Distributed by Syracuse University (1931).

ON THE INDEPENDENCE OF CERTAIN ESTIMATES OF VARIANCE[1]

By Allen T. Craig

1. **Introduction.** It is well known that a necessary and sufficient condition that several statistics be independent in the probability sense, is that the characteristic function of the joint distribution of these statistics shall equal identically the product of the characteristic functions of the distributions of the individual statistics. Thus, if $x_1, x_2, \cdots, x_N$ are N independently observed values of a variable x which is subject to the distribution function $f(x)$, and if $\theta_1, \theta_2, \cdots, \theta_s$ are s statistics, each computed from the N observed values of x, the characteristic function of the joint distribution of the s statistics is given by

$$\varphi(t_1, t_2, \cdots, t_s) = \int \cdots \int e^{it_1\theta_1 + \cdots + it_s\theta_s} f(x_1) \cdots f(x_N)\, dx_N \cdots dx_1.$$

Here, $i = \sqrt{-1}$ and the limits of integration are taken so as to include all admissible values of x. Since the characteristic function of the distribution of θ_v, $v = 1, 2, \cdots, s$, is given by

$$\varphi_v(t_v) = \int \cdots \int e^{it_v\theta_v} f(x_1) \cdots f(x_N)\, dx_N \cdots dx_1,$$

the necessary and sufficient condition for the independence of the s statistics can be written

$$\varphi(t_1, \cdots, t_s) = \varphi_1(t_1) \cdots \varphi_s(t_s), \tag{1}$$

for all real values of $t_1, t_2, \cdots t_s$.

An important phase of sampling theory in statistics is that in which the variable x is subject to the normal distribution function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \qquad -\infty \leq x \leq \infty,$$

and $\theta_1, \cdots, \theta_s$ are s real symmetric quadratic forms in the N independently observed values of x. That is,

$$\theta_1 = \sum_{j=1}^{N} \sum_{k=1}^{N} a_{jk} x_j x_k,$$

$$\theta_2 = \sum_{j=1}^{N} \sum_{k=1}^{N} b_{jk} x_j x_k,$$

$$\vdots$$

$$\theta_s = \sum_{j=1}^{N} \sum_{k=1}^{N} p_{jk} x_j x_k,$$

[1] Presented to the Institute of Mathematical Statistics on December 30, 1937, at the invitation of the progam committee. In the paper, we discuss, from a slightly different point of view, some of the material found in the references given at the close of the paper.

so that

$$\varphi(t_1, \cdots, t_s) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{iT}\, dx_N \cdots dx_1, \tag{2}$$

where $T = t_1 \sum \sum a_{jk}x_j x_k + \cdots + t_s \sum \sum p_{jk} x_j x_k - \frac{1}{2i\sigma^2} \sum x_j^2$. If $A_1, \cdots, A_s$ denote the real symmetric matrices of the s quadratic forms, the characteristic function can be written

$$\varphi(t_1, \cdots, t_s) = | I - 2i\sigma^2 t_1 A_1 - \cdots - 2i\sigma^2 t_s A_s |^{-\frac{1}{2}},$$

where I is the unit matrix of order N and the vertical bars indicate the determinant of the matrix within them. Similarly, the characteristic function of the distribution of θ_v is given by

$$\varphi_v(t_v) = | I - 2i\sigma^2 t_v A_v |^{-\frac{1}{2}},$$

so that a necessary and sufficient condition for the independence of the s real symmetric quadratic forms can be written

$$| I - 2i\sigma^2 t_1 A_1 - \cdots - 2i\sigma^2 t_s A_s | = \prod_{v=1}^{s} | I - 2i\sigma^2 t_v A_v |, \tag{3}$$

for all real values of $t_1, t_2, \cdots, t_s$.

Although equation (3) is fundamental and is of considerable value in certain problems, it should be remarked that it is frequently rather tedious to use. This suggests that by strengthening the hypotheses, it may be possible to establish another necessary and sufficient condition which, in certain cases, may be easier to use.

2. Certain quadratic forms. In order to lead up to such a theorem as that suggested at the close of the last section, we first consider two theorems regarding real symmetric matrices.

Theorem I. *Let $A_1, A_2, \cdots, A_s$ be s real symmetric matrices, each of order N, such that $A_1 + A_2 + \cdots + A_s = I$, where I is the unit matrix of order N. Let r_v, $v = 1, 2, \cdots, s$, be respectively the ranks of the matrices A_v. If $r_1 + r_2 + \cdots + r_s = N$, each of the non-zero roots of the characteristic equations*[2] *of the matrices A_v is $+1$.*

If $s = 2$, the theorem is almost self-evident. For the characteristic equation of A_2 is $| A_2 - \lambda I | = 0$, which, since $A_1 + A_2 = I$, can be written $| I - A_1 -$

---

[2] By the characteristic equation of the square matrix A is meant the algebraic equation of degree N in λ, $| A - \lambda I | = 0$. If A is real and symmetric and the rank of A is r, the characteristic equation has exactly r real non-zero roots and $N - r$ zero roots. Cf. Kowalewski, Einführung in die Determinanten-Theorie (1909) pp. 126–128.

$\lambda I \mid = 0$ or $\mid A_1 - (1 - \lambda)I \mid = 0$. But the last equation is the characteristic equation of A_1 with λ replaced by $1 - \lambda$. Thus the roots of the equation $\mid A_1 - \lambda I \mid = 0$ are one minus the roots of $\mid A_2 - \lambda I \mid = 0$. Since the equation $\mid A_2 - \lambda I \mid = 0$ has $N - r_2$ zero roots, the equation $\mid A_1 - \lambda I \mid = 0$ has $N - r_2$ roots equal to $+1$. But $r_1 = N - r_2$ so that all the non-zero roots of $\mid A_1 - \lambda I \mid = 0$ are $+1$. A similar statement holds for the roots of $\mid A_2 - \lambda I \mid = 0$.

In general, we have $A_1 + A_2 + \cdots + A_s = I$ and $r_1 + r_2 + \cdots + r_s = N$. Let $B_1 = A_2 + A_3 + \cdots + A_s$ and denote by R_1 the rank of B_1. Thus[3] $R_1 \leq r_2 + r_3 + \cdots + r_s$. Now $A_1 + B_1 = I$ and the equation $\mid A_1 - \lambda I \mid = 0$ has exactly $N - r_1$ zero roots. Since the roots of $\mid B_1 - \lambda I \mid = 0$ are one minus the roots of $\mid A_1 - \lambda I \mid = 0$, the first of these two equations has at least $N - r_1$ non-zero roots so that $R_1 \geq N - r_1 = r_2 + r_3 + \cdots + r_s$. From $r_2 + r_3 + \cdots + r_s \leq R_1 \leq r_2 + r_3 + \cdots + r_s$ we deduce the equality so that the argument in the case of $s = 2$ applies to the matrices A_1 and B_1. In particular, then, each of the non-zero roots of $\mid A_1 - \lambda I \mid = 0$ is $+1$. By writing $B_2 = A_1 + A_3 + \cdots + A_s$, $B_3 = A_1 + A_2 + A_4 + \cdots + A_s$, and so on, and repeating the argument in each instance, we see that the theorem holds.

Theorem II. *Let $A_1, A_2, \cdots, A_s$ be s real symmetric matrices which satisfy the conditions of Theorem* I. *There then exist $s - 1$ real orthogonal matrices of order N, say $L_1, L_2, \cdots, L_{s-1}$, such that each of the s matrices*

$$L'_{s-1} \cdots L'_1 A_v L_1 \cdots L_{s-1}, \qquad v = 1, 2, \cdots, s,$$

is a diagonal matrix[4] *with the r_v non-zero elements on the principal diagonal equal to $+1$. Necessarily, the sum of these s matrices is the identity matrix.*

In proof of the theorem we shall, to save space, restrict ourselves to the case of $s = 3$, although the method we use will be readily seen to be entirely general. Since A_1 is real and symmetric and since, by Theorem I, the r_1 non-zero roots of the characteristic equation of A_1 are $+1$, there exists a real orthogonal matrix of order N, say L_1, such that

$$L'_1 A_1 L_1 = \left\|\begin{array}{cccc|ccc} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \vdots & & \vdots \\ \hline 0 & \cdots & \cdots & \cdots & 0 & \cdots & 0 \\ \vdots & & & & \vdots & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \cdots & 0 \end{array}\right\|$$

where L'_1 is the conjugate of L_1 and where, merely as a convenience of notation, we have placed the r_1 non-vanishing elements of the principal diagonal in the first r_1 rows and columns. If then, in both members of the equation $A_1 + A_2 + A_3 = I$, we multiply on the left by L'_1 and on the right by L_1, we have

[3] Cf. Bôcher, Introduction to Higher Algebra (1921) p. 62.

[4] By a diagonal matrix we mean a matrix whose elements not on the principal diagonal are zero.

$$(4) \qquad \left\| \begin{array}{cccc|ccc} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ \hline 0 & \cdots & \cdots & \cdots & 0 & \cdots & 0 \\ \vdots & & & & \vdots & & \vdots \\ 0 & & & & 0 & & 0 \end{array} \right\| + L_1'A_2L_1 + L_1'A_3L_1 = I,$$

since $L_1'IL_1 = IL_1'L_1 = I$. The matrices $L_1'A_2L_1$ and $L_1'A_3L_1$ are real, symmetric, and the ranks are r_2 and r_3, since L_1 is non-singular. Moreover, the non-zero roots of the characteristic equations of the two matrices are $+1$; for $|L_1'A_2L_1 - \lambda I| = |L_1'(A_2 - \lambda I)L_1| = |L_1'||A_2 - \lambda I||L_1|$, and similarly for the matrix $L_1'A_3L_1$. Now if a real symmetric matrix is positive definite, that is, if all the non-zero roots of its characteristic equation are positive, then[5] all the elements on the principal diagonal are positive or zero, and, if an element on the principal diagonal is zero, all the elements in the row and column in which that element lies are zero. These two facts regarding a real symmetric positive definite matrix, in conjunction with equation (4), require that the matrices $L_1'A_2L_1$ and $L_1'A_3L_1$ be of the forms

$$\left\| \begin{array}{ccc|ccc} 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & b_{r_1+1,r_1+1} & \cdots & b_{r_1+1,N} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & b_{N,r_1+1} & & b_{NN} \end{array} \right\| \quad \text{and} \quad \left\| \begin{array}{ccc|ccc} 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & c_{r_1+1,r_1+1} & \cdots & c_{r_1+1,N} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & c_{N,r_1+1} & \cdots & c_{NN} \end{array} \right\|$$

respectively. Now the real symmetric matrix

$$C = \left\| \begin{array}{ccc} b_{r_1+1,r_1+1} & \cdots & b_{r_1+1,N} \\ \vdots & & \vdots \\ b_{N,r_1+1} & \cdots & b_{NN} \end{array} \right\|$$

is of order $N - r_1$, its rank is r_2, and its characteristic equation has r_2 roots equal to $+1$. There then exists a real orthogonal matrix M of order $N - r_1$, say

$$M = \left\| \begin{array}{ccc} m_{r_1+1,r_1+1} & \cdots & m_{r_1+1,N} \\ m_{N,r_1+1} & \cdots & m_{NN} \end{array} \right\|,$$

such that

$$M'CM = \left\| \begin{array}{cccc|ccc} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \vdots & & \\ \vdots & \vdots & & \vdots & \vdots & & \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ \hline 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & & & & \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{array} \right\|.$$

[5] Cf. Cullis, Matrices and Determinoids (1918) vol. 2, p. 302.

Again, to simplify the notation, we have placed the r_2 non-vanishing elements of the principal diagonal in the first r_2 rows and columns. Consider the orthogonal matrix of order N

$$L_2 = \left\| \begin{array}{cccc|ccc} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & & & \\ \vdots & & & & & & \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ \hline 0 & \cdots & \cdots & 0 & m_{r_1+1,r_1+1} & \cdots & m_{r_1+1,N} \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & m_{N,r_1+1} & \cdots & m_{NN} \end{array} \right\|.$$

It is evident that $L_2'(L_1'A_1L_1)L_2 = L_1'A_1L_1$. If then, both members of $L_1'A_1L_1 + L_1'A_2L_1 + L_1'A_3L_1 = I$ are multiplied on the left by L_2' and on the right by L_2, we get

$$\left\| \begin{array}{ccc|ccc} 1 & 0 \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & & & \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{array} \right\| + \left\| \begin{array}{ccc|ccc|ccc} 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & & & \vdots & & \\ 0 & \cdots & 0 & & & & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 1 & 0 \cdots & 0 & & & \vdots \\ \vdots & & \vdots & \vdots & & \vdots & & & \vdots \\ \vdots & & \vdots & 0 & \cdots & 1 & & & \vdots \\ \hline \vdots & & \vdots & & & & & & \vdots \\ 0 & & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{array} \right\|$$

$$+ \left\| \begin{array}{ccc|ccc} 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & d_{r_1+1,r_1+1} & \cdots & d_{r_1+1,N} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & d_{N,r_1+1}, & \cdots & d_{NN} \end{array} \right\| = I.$$

From this last equation, it follows that $d_{jk} = 0$, $j \neq k$, $d_{jj} = 0$, $j = r_1 + 1, \cdots, r_1 + r_2$ and $d_{jj} = 1$, $j = r_1 + r_2 + 1, \cdots, N$. The third matrix in the left member of preceding equation then takes the form

$$\left\| \begin{array}{ccc|ccc|ccc} 0 & \cdots & 0 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & & & & & \vdots \\ 0 & \cdots & 0 & & & & & & \vdots \\ \hline 0 & & & 0 & \cdots & 0 & & & \vdots \\ \vdots & & & \vdots & & \vdots & & & \vdots \\ \vdots & & & 0 & \cdots & 0 & & & 0 \\ \hline \vdots & & & & & & 1 & 0 \cdots & 0 \\ \vdots & & & & & & \vdots & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 1 \end{array} \right\|.$$

This establishes Theorem II when $s = 3$. The procedure may be continued in a fairly obvious manner so as to justify the theorem for any finite positive integer s.

With the aid of Theorems I and II, we are now able to state and prove a very useful theorem on the independence of certain quadratic forms of normally and independently distributed variables. The theorem follows.

Theorem III. *Let $x_1, x_2, \cdots, x_N$ be N independent values of a normally distributed variable x and let $\theta_1, \cdots, \theta_s$ be s real symmetric quadratic forms in these N variables, where $\sum_1^s \theta_i = \sum_1^N x_j^2$. If $r_1, r_2, \cdots, r_s$ denote respectively the ranks of the quadratic forms, a necessary and sufficient condition that the s forms be independent in the probability sense is that $r_1 + r_2 + \cdots + r_s = N$.*

Consider the characteristic function of the joint distribution of the s forms as given by equation (2). In accordance with Theorem II, we can successively introduce new variables by performing real linear transformations with orthogonal matrices $L_1, L_2, \cdots, L_{s-1}$ respectively in such a way that[6] T becomes

$$T = t_1 \sum_1^{r_1} y_j^2 + t_2 \sum_{r_1+1}^{r_1+r_2} y_j^2 + \cdots + t_s \sum_{r_1+\cdots+r_{s-1}+1}^{N} y_j^2 - \frac{1}{2i\sigma^2} \sum_1^N y_j^2.$$

Since each transformation is orthogonal, the absolute value of the Jacobian in each instance is unity. Thus the right member of (2) can now be written as the product of s sets of integrals, the sets containing $r_1, r_2, \cdots, r_s$ integrals respectively. That is,

$$\varphi(t_1, \cdots, t_s) = \varphi_1(t_1) \cdots \varphi_s(t_s),$$

which is equation (1). Hence the theorem.

Under the conditions of Theorem III, the characteristic function of the distribution of θ_v is found by direct integration to be

$$\varphi_v(t_v) = (1 - 2i\sigma^2 t_v)^{-\frac{r_v}{2}}.$$

[6] If the variables in a symmetric quadratic form with matrix A are transformed by a linear transformation with matrix B, the new form has the matrix $B'AB$. Cf. Bôcher, p. 129. It should be remarked that these $s - 1$ successive orthogonal transformations can be combined into a single orthogonal transformation with matrix $L = L_1 L_2 \cdots L_{s-1}$. For if, by means of a linear transformation with matrix L_1, we pass from the variables $x_1, \cdots, x_N$ to the variables $x_1', \cdots, x_N'$, in which the old variables are expressed explicitly in terms of the new, and thence to variables $x_1'', \cdots, x_N''$ by means of a linear transformation with matrix L_2, the transformation with matrix $L_1 L_2$ will carry us directly from the x's to the x'''s. This extends to any finite number of transformations. Since the product of any two orthogonal matrices is an orthogonal matrix (and hence the product of a finite number of them), we see that the remark is justified. Cf. Bôcher, p. 68 and Kowalewski, p. 161. Note that Bôcher expresses the new variables explicitly in terms of the old.

Thus,

$$f_v(\theta_v)\,d\theta_v = d\theta_v\,\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it_v\theta_v}\,\varphi_v(t_v)\,dt_v$$

$$= \frac{1}{2^{\frac{r_v}{2}}\,\Gamma\left(\frac{r_v}{2}\right)}\left(\frac{\theta_v}{\sigma^2}\right)^{\frac{r_v}{2}-1} e^{-\frac{1}{2}\left(\frac{\theta_v}{\sigma^2}\right)}\,\frac{d\theta_v}{\sigma^2},$$

so that the variables θ_v/σ^2 are distributed in accordance with Chi-square distributions with r_v degrees of freedom.[7] Accordingly, when the conditions of Theorem III are satisfied, we may deduce not merely the mutual independence of the θ_v but also the nature of their distributions.

3. **Applications to the analysis of variance.** In the analysis of variance, $N = ab$ independently observed values of a normally distributed variable are classified into a rows and b columns in accordance with some relevant scheme:

$$\begin{array}{l} x_{11}\,,\ x_{12}\,,\ \cdots\,,\ x_{1b} \\ x_{21}\,,\ x_{22}\,,\ \cdots\,,\ x_{2b} \\ \ \vdots \qquad \vdots \qquad\quad \vdots \\ x_{a1}\,,\ x_{a2}\,,\ \cdots\,,\ x_{ab}\,. \end{array}$$

With the notation $\bar{x}_{j\cdot}$, $\bar{x}_{\cdot k}$, $\bar{x}$ to denote respectively the arithmetic mean of the jth row, the kth column, and the entire set, it is readily seen that

$$(5)\quad \begin{aligned} \sum_1^a\sum_1^b (x_{jk}-\bar{x})^2 &= b\sum_1^a(\bar{x}_{j\cdot}-\bar{x})^2 + a\sum_1^b(\bar{x}_{\cdot k}-\bar{x})^2 + \sum_1^a\sum_1^b (x_{jk}-\bar{x}_{j\cdot}-\bar{x}_{\cdot k}+\bar{x})^2 \\ &= \theta_1 + \theta_2 + \theta_3 \end{aligned}$$

is an identity in the $N = ab$ values of x. It is quite straightforward to exhibit each of the three terms in the right member of (5) as a real symmetric quadratic form in the N variables x_{jk} and to show that the ranks are $r_1 = a - 1$, $r_2 = b - 1$, $r_3 = (a-1)(b-1)$. By the device of adding $\theta_4 = \frac{1}{ab}\left(\sum\sum x_{jk}\right)^2 = N\bar{x}^2$ to both members of (5), we have $\sum\sum x_{jk}^2 = \theta_1+\theta_2+\theta_3+\theta_4$. Moreover, the rank of θ_4 is $r_4 = 1$. Thus $r_1 + r_2 + r_3 + r_4 = ab = N$ and, by Theorem III, we see that the four quadratic forms are mutually independent. In particular, θ_1, θ_2 and θ_3 are independent, and each, measured in units of σ^2, is distributed as is Chi-square with its appropriate number of degrees of freedom.

THE UNIVERSITY OF IOWA.

[7] By the number of degrees of freedom of a real symmetric quadratic form of normally and independently distributed variables, we mean the rank of the matrix of the form.

REFERENCES

(1) M. S. Bartlett and J. Wishart, *The distribution of second order moment statistics in a normal system.* Proceedings of the Cambridge Philosophical Society, vol. 28 (1931–32) pp. 455–459.

The generalized product moment distribution in a normal system. Same journal, vol. 29 (1932–33) pp. 260–270.

(2) W. G. Cochran, The distribution of quadratic forms in a normal system. Proceedings of the Cambridge Philosophical Society, vol. 30 (1933–34) pp. 178–191.

(3) R. A Fisher, Applications of Student's distribution. Metron, vol. 5 (1925) pp. 90–104.

Statistical Methods for Research Workers (1934) pp. 210–272.

(4) S. S. Wilks, Statistical Inference (1936–37) pp. 38, 44–45.

VARIANCE OF A GENERAL MATCHING PROBLEM*

By Joseph A. Greenwood

Let us match two decks of cards: (A) composed of t distinct groups of s identical symbols each, and (B) a target deck composed of i_1 symbols of the first kind, i_2 of the second, etc., such that

$$i_1 + i_2 + \cdots + i_t = st = n. \tag{1}$$

It is not necessary that all the i's be different from zero.

(a) *Forming the Relative Frequency Table.* The first part of the paper is concerned with forming a 2x2-way table showing the relative frequencies of hits and misses of all pairs of cards in the target deck. The notation $\overset{i}{0}$ indicates a miss at the ith card of the target deck, $\overset{i}{1}$ a hit. $\overset{i}{0} = j$ indicates a miss at the ith card, with the matching card identical to the jth target card.

Case I. *ith and jth target cards the same symbol.*

i	j	Theoretical freq.	Weighted freq.	
If 0 then —[1]	0	$n - s - 1$	$(t-1)(n-s-1)$	2.1
0	1	s	$(t-1)\,s = n - s$	2.2
1	0	$n - s$	$n - s$	2.3
1	1	$s - 1$	$s - 1$	2.4
			Total $= t(n-1)$	

(2)

But $\overset{i}{0}$ occurs in $(t-1)/t$ of the events. Thus we must weight 2.1 and 2.2 with a factor $(t-1)$, giving the last column in (2).

Case II. *ith and jth target cards different*

i	j	Theoretical freq.	Weighted freq.	
If $0 = j$ then —	0	$n - s$	$n - s$	3.1
$0 = j$	1	$s - 1$	$s - 1$	3.2
$0 \neq j$	0	$n - s - 1$	$(n-s-1)(t-2)$	3.3
$0 \neq j$	1	s	$s(t-2)$	3.4
1	0	$n - s - 1$	$n - s - 1$	3.5
1	1	s	s	3.6
			Total $= t(n-1)$	

(3)

* Presented to the American Mathematical Society, September 9, 1937.

[1] Read, 'then out of $n - 1$ times'.

But $\overset{i}{0} = j$ occurs in $1/(t-1)$ of all events $\overset{i}{0}$, and $\overset{i}{1}$ occurs in $1/t$ of all events $\overset{i}{1} + \overset{i}{0}$. Therefore entries 3.3 and 3.4 must be weighted with the factor $(t-2)$, and then entries 3.1, 3.2, 3.3 and 3.4 must be weighted with the factor $(t-1)$. It is important that the totals of the two parts to be weighted be equal before the weighting factors are applied. This gives rise to the last column in table (3).

Now the number of ways the ith card can be like the jth card of the target deck is[2]

$$\alpha_1 = \sum_{f=1}^{t} \binom{i_f}{2}$$

The number of ways they can be unequal is

$$\alpha_2 = \sum_{u<v}^{1,\dots,t} (i_u i_v) = \binom{n}{2} - \alpha_1. \tag{4}$$

Since the totals of the last columns of the two tables are equal we weight the entries of their last columns with α_1 and α_2, respectively. So, combining 3.1, 3.3 and 3.2, 3.4 we form α_1 times (2) + α_2 times (3) to give the new table

i	j	Relative frequencies	
0	0	$(n-s-1)(t-1)\alpha_1 + [(t-1)(n-s-1)+1]\alpha_2$	
0	1	$(n-s)\alpha_1 + (n-s-1)\alpha_2$	
1	0	$(n-s)\alpha_1 + (n-s-1)\alpha_2$	(5)
1	1	$(s-1)\alpha_1 + s\alpha_2$.	

Now using the entries from (5) form the 2x2-way table

		Total	
$(t-1)(n-s-1)\alpha_1 + [(t-1)(n-s-1)+1]\,\alpha_2$	$(n-s)\alpha_1 + (n-s-1)\alpha_2$	$(tn-n-t+1)(\alpha_1+\alpha_2)$	
$(n-s)\alpha_1 + (n-s-1)\alpha_2$	$(s-1)\alpha_1 + s\alpha_2$	$(n-1)(\alpha_1+\alpha_2)$	(6)
$(tn-n-t+1)(\alpha_1+\alpha_2)$	$(n-1)(\alpha_1+\alpha_2)$	$t(n-1)(\alpha_1+\alpha_2)$	

---

[2] If $i_v < 2$ define $\binom{i_v}{2} = 0$.

(b) *Obtaining the Correlation, Variance and Maximal Conditions.* Substituting from table (6) into the formulas given by Yule[3] for δ and the coefficient of correlation r, we obtain the average correlation

$$r = \frac{\alpha_2 + (1-t)\alpha_1}{(tn - n - t + 1)(\alpha_1 + \alpha_2)} = \frac{\binom{n}{2} - t\alpha_1}{\binom{n}{2}(tn - n - t + 1)}$$

$$= \frac{t\alpha_2 + (1-t)\binom{n}{2}}{\binom{n}{2}(tn - n - t + 1)} \tag{7}$$

by (4).

We now give a proof that r is a maximum when $i_j = s$ $(j = 1, \cdots, t)$. From (7) it is sufficient to show that under the same conditions α_2 is a maximum.

Let $i_j = s + \delta_j$, then

$$\sum^{1,\cdots,t} \delta_j = 0 \qquad \text{by (1).} \tag{8}$$

$$\alpha_2 = \sum_{u<v}^{1,\cdots,t} (s + \delta_u)(s + \delta_v) = \binom{t}{2} s^2 + \sum_{u<v}^{1,\cdots,t} \delta_u \delta_v \quad \text{by (8).}$$

Assume some $\delta_u \neq 0$ and

$$\sum_{u<v}^{1,\cdots,t} \delta_u \delta_v \geqq 0. \tag{9}$$

Add

$$\sum^{1,\cdots,t} \delta_u^2 + \sum_{u<v}^{1,\cdots,t} \delta_u \delta_v$$

to both sides of (9). Then

$$\left(\sum^{1,\cdots,t} \delta_u\right)^2 \geqq \sum^{1,\cdots,t} \delta_u^2 + \sum_{u<v}^{1,\cdots,t} \delta_u \delta_v \tag{10}$$

or $0 \geqq$ a positive number. This necessarily implies the desired result.

---

[3] Yule, G. U. *An Introduction to the Theory of Statistics*, London: Griffin and Co., 1927, pp. 216-217. The table can be symbolized with

				Total
	a_1	a_2	a_3	
	b_1	b_2	b_3	
Total	c_1	c_2	c_3	

$$\delta = b_2 - (c_2 b_3 / c_3)$$

He then gives $r = \delta c_3 / \sqrt{c_1 c_2 a_3 b_3}$, the correlation coefficient.

Yule[4] gives an expression for the variance in a situation which includes the present problem as a special case, to be

$$\sigma^2 = npq[1 + r(n - 1)], \tag{11}$$

where r is the average correlation between all pairs of variables. Substituting our result (7) in (11) with $p = 1/t$ gives the desired variance.

It is interesting to note that when $i_j = s$, $(j = 1, \cdots, t)$ r reduces to $1/(n-1)^2$ giving

$$\sigma^2 = \frac{n^2(t-1)}{t^2(n-1)} = \frac{n}{n-1}\,\sigma_b^2$$

where σ_b^2 is the variance of the binomial case.[5]

DUKE UNIVERSITY.

---

[4] Op. cit., p. 286.

[5] Concerning this special case see also Bartlett, M. S. *Properties of sufficiency and statistical tests.* Proc. Royal Soc. A. 1937, CLX, 268–282.

Olds, E. G. *A moment-generating function useful in certain matching problems.* Abstract No. 428, Bull. Amer. Math. Soc. 1937, XLIII, 779.

THE LARGE-SAMPLE DISTRIBUTION OF THE LIKELIHOOD RATIO FOR TESTING COMPOSITE HYPOTHESES[1]

By S. S. Wilks

By applying the principle of maximum likelihood, J. Neyman and E. S. Pearson[2] have suggested a method for obtaining functions of observations for testing what are called *composite statistical hypotheses*, or simply *composite hypotheses*. The procedure is essentially as follows: A population K is assumed in which a variate x (x may be a vector with each component representing a variate) has a distribution function $f(x, \theta_1, \theta_2, \cdots \theta_h)$, which depends on the parameters $\theta_1, \theta_2 \cdots \theta_h$. A *simple hypothesis* is one in which the θ's have specified values. A set Ω of admissible hypotheses is considered which consists of a set of simple hypotheses. Geometrically, Ω may be represented as a region in the h-dimensional space of the θ's. A set ω of simple hypotheses is specified by taking all simple hypotheses of the set Ω for which $\theta_i = \theta_{0i}$, $i = m + 1, m + 2, \cdots h$.

A random sample O_n of n individuals is considered from K. O_n may be geometrically represented as a point in an n-dimensional space of the x's. The probability density function associated with O_n is

$$P = \prod_{\alpha=1}^{n} f(x_\alpha, \theta_1, \theta_2, \cdots \theta_h) \tag{1}$$

Let $P_\Omega(O_n)$ be the least upper bound of P for the simple hypotheses in Ω, and $P_\omega(O_n)$ the least upper bound of P for those in ω. Then

$$\lambda = \frac{P_\omega(O_n)}{P_\Omega(O_n)} \tag{2}$$

is defined as the likelihood ratio for testing the composite hypothesis H that O_n is from a population with a distribution characterized by values of the θ_i for some simple hypothesis in the set ω. When we say that H is true, we shall mean that O_n is from some population of the set just described. In most of the cases of any practical importance, P and its first and second derivatives with respect to the θ_i are continuous functions of the θ_i almost everywhere in a certain region of the θ-space for almost all possible samples O_n. We shall only consider the case in which $P_\Omega(O_n)$ and $P_\omega(O_n)$ can be determined from the first and second order derivatives with respect to the θ's.

---

[1] Presented to the American Mathmatical Society, March 26, 1937.

[2] Phil. Trans. Roy. Soc. London, Ser. A, Vol. 231, p. 295.

A considerable number of currently used statistical functions for making tests of significance can be expressed in terms of λ ratios, and in many cases involving normal distribution theory, the exact sampling distribution of λ is known. However, it is often useful when dealing with large samples to have an approximation to the distribution of λ. We shall consider such an approximation for those cases (which include most of the ones of any practical importance) in which *optimum* estimates of the θ's exist. That is, we shall assume the existence of functions $\bar{\theta}_i(x_1, \cdots x_n)$(maximum likelihood estimates of the θ_i) such that[3] their distribution is

$$(3) \qquad \frac{|c_{ij}|^{\frac{1}{2}}}{(2\pi)^{h/2}} e^{-\frac{1}{2}\sum_{i,j=1}^{h} c_{ij}z_iz_j} (1+\phi)\, dz_1 \cdots dz_h$$

where $z_i = (\bar{\theta}_i - \theta_i)\sqrt{n}$, $c_{ij} = -E\left(\dfrac{\partial^2 \log f}{\partial\theta_i \partial\theta_j}\right)$, E denoting mathematical expectation, and ϕ is of order $1/\sqrt{n}$ and $||c_{ij}||$ is positive definite. Denoting (3) by $J dz_1 dz_2 \cdots dz_h$, and differentiating J with respect to θ_k, we get

$$(4) \qquad \frac{1}{2}\left(\frac{1}{|c_{ij}|}\frac{\partial |c_{ij}|}{\partial\theta_k} - \sum_{i,j}\frac{\partial c_{ij}}{\partial\theta_k} z_iz_j + \sqrt{n}\sum_j c_{kj}z_j\right)J, \quad k = 1, 2, \cdots h$$

Since $c_{ij} = O(1)$ and $|c_{ij}| \neq 0$, it can be seen from (4) that the values of θ_k which maximize J differ from $\bar{\theta}_k$, $k = 1, 2, \cdots h$, by terms of order $1/\sqrt{n}$. Therefore, the maximum $P_\Omega(O_n)$ of J with respect to the θ_k is $\dfrac{|c_{ij}|^{\frac{1}{2}}}{(2\pi)^{k/2}}(1+\phi')$, where $\phi' = O(1/\sqrt{n})$.

To get $P_\omega(O_n)$, we let $\theta_i = \theta_{0i}$, $i = m+1, m+2, \cdots h$, and note that J can be written as

$$(5) \qquad J_0 = \frac{|c_{0ij}|^{\frac{1}{2}}}{(2\pi)^{h/2}} e^{-\frac{1}{2}\sum_{i,j=1}^{m} c_{ij}z_i'z_j' - \frac{1}{2}\chi_0^2} (1+\phi_0')$$

where

$$(6) \qquad \chi_0^2 = \sum_{i,j=m+1}^{h} c_{ij}' z_iz_j, \qquad \phi_0' = O(1/\sqrt{n})$$

and $||c_{ij}'||$ is the inverse of the matrix obtained by deleting the first m rows and first m columns from $||c_{ij}||^{-1}$ and $z_i' = z_i - L_i$, L_i being a linear function of $\theta_{0,m+1} \cdots \theta_{0k}$, and c_{0ij} is the value of c_{ij} with $\theta_i = \theta_{0i}$, $i = m+1, m+2, \cdots h$, that is, when H is true. Taking the maximum $P_\omega(O_n)$ of expression (5) with respect to $\theta_1, \theta_2, \cdots \theta_m$, we get

$$(7) \qquad P_\omega = \frac{|c_{0ij}|^{\frac{1}{2}}}{(2\pi)^{h/2}} e^{-\frac{1}{2}\chi_0^2}(1+\phi_0'') \qquad \phi_0'' = O(1/\sqrt{n})$$

[3] For conditions under which the $\bar{\theta}$'s exist which are distributed according to (3), see J. L. Doob, Probability and Statistics, Trans. Amer. Math. Soc. Vol. 36, p. 759-775.

Hence, when H is true, we have, from (5) and (7)

$$\lambda = \frac{P_\omega(O_n)}{P_\Omega(O_n)} = e^{-\frac{1}{2}\chi_0^2}(1 + O(1/\sqrt{n})). \tag{8}$$

Therefore, except for terms of order $1/\sqrt{n}$,

$$-2 \log \lambda = \chi_0^2 . \tag{9}$$

Now, the characteristic function of $-2 \log \lambda$ is

$$\begin{aligned} \phi(t) &= E(e^{it(-2\log\lambda)}) = \int \cdots \int J_0 e^{it(\chi_0^2 + O(1/\sqrt{n}))}\, dz_1 \cdots dz_h \\ &= \frac{|c_{ij}|^{\frac{1}{2}}}{(2\pi)^{h/2}} \int \cdots \int e^{-\frac{1}{2}\sum_{i,j=1}^{m} c_{ij}z'_i z'_j + \chi_0^2(it - \frac{1}{2})} (1 + O(1/\sqrt{n}))\, dz_1 \cdots dz_h . \end{aligned} \tag{10}$$

It can be shown that on any finite interval $|t| < a$, $\phi(t)$ approaches uniformly, as $n \to \infty$, the function

$$(\tfrac{1}{2})^{\frac{h-m}{2}}(\tfrac{1}{2} - it)^{-\frac{h-m}{2}}. \tag{11}$$

But (11) is the characteristic function of any quantity distributed like χ^2 with $h - m$ degrees of freedom.

We can summarize in the

Theorem: If a population with a variate x is distributed according to the probability function $f(x, \theta_1, \theta_2 \cdots \theta_h)$, such that optimum estimates $\bar{\theta}_i$ of the θ_i exist which are distributed in large samples according to (3), then when the hypothesis H is true that $\theta_i = \theta_{0i}$, $i = m + 1, m + 2, \cdots h$, the distribution of $-2 \log \lambda$, where λ is given by (2) is, except for terms of order $1/\sqrt{n}$, distributed like χ^2 with $h - m$ degrees of freedom.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

ON DIFFERENTIAL OPERATORS DEVELOPED BY O'TOOLE

BY M. ZIAUD-DIN

1. O'Toole in his paper 'Symmetric Functions and Symmetric Functions of Symmetric Functions' [Ann. Statist. 2. (1931)102–49], has expressed Monomial Symmetric Functions $\sum a^{p_1} b^{p_2} c^{p_3} \cdots$, in terms of power-sums, s_r.

The Monomial Symmetric Functions can be written in partition notation as $(p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots)$ where $k_1, k_2, \cdots$ denote the repetitions of parts.

To express $(p_1^{k_1} p_2^{k_2} p_3^{k_3} \cdots)$ as a function of s_r', O'Toole has developed operators d_r and D_r, connected by the formulae,[1]

$$d_r = \frac{d}{ds_r},$$

$$\text{(A)} \qquad rd_r = \frac{(-1)^{r+1} \sum (-1)^{r+k}(k-1)!\, r \cdot D_A^{k_1} D_B^{k_2} \cdots}{k_1! \, k_2! \cdots}$$

$$\text{(B)} \qquad r! D_r = \frac{\sum r! d_A^{k_1} d_B^{k_2} \cdots}{k_1! \, k_2! \cdots},$$

$$\text{where } k_1 A + k_2 B + \cdots = r$$

$$k_1 + k_2 + \cdots = k.$$

In this paper it will be shown that these operational relations are easily deduced from the operators d_r and D_r of Hammond, used for expressing Monomial Symmetric Functions as functions of Elementary Symmetric Functions, a_r.

For the sake of distinction I shall use q_r and Q_r for the operators employed by O'Toole and keep d_r and D_r for Hammond's Operators.

Macmahon has dealt with Hammond's operators in his Combinatory Analysis Vol. I. Cambridge University Press (1915), where they are defined[2] as

$$D_r = \frac{1}{r!}(d_1^r) \qquad \text{and} \qquad d_r = \frac{d}{da_r} + a_1 \frac{d}{da_{r+1}} + a_2 \frac{d}{da_{r+2}} + \cdots, \qquad (1).$$

2. It is known[3] that

$$\log(1 - a_1 x + a_2 x^2 - a_3 x^3 + \cdots) = -\left(s_1 x + \frac{1}{2} s_2 x^2 + \cdots + \frac{1}{r} s_r x^r + \cdots\right).$$

---

[1] O'Toole, Loc. cit., p. 120.

[2] Macmahon. Comb. Analysis. I. 27–28.

[3] Ibid., p. 6.

Now operate on the right hand side with d_r, and with its equivalent in (1) on the left hand side. Equating coefficients of x^r on both sides, we obtain,

$$d_r s_r = (-1)^{r-1} r; \qquad d_r s_k = 0 \text{ when } r \neq k,$$

which yields

$$d_r = (-1)^{r-1} r \frac{d}{ds_r} = (-1)^{r-1} r q_r. \tag{2}$$

The operator Q_r exactly behaves like D_r. From the formula[4]

$$d_r - D_1 d_{r-1} + D_2 d_{r-2} - D_3 d_{r-3} + \cdots + (-1)^r r d_r = 0,$$

which is in complete correspondence with Newton's recurrence relation, we derive

$$\begin{aligned} d_1 &= D_1 \\ d_2 &= D_1^2 - 2D_2 \\ d_3 &= D_1^3 - 3D_1 D_2 + 3D_3. \\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{C}$$

By multinomial theorem

$$d_r = \frac{\sum (-1)^{r+k}(k-1)! r D_A^{k_1} D_B^{k_2} \cdots}{k_1! k_2! \cdots}$$

using (2) we at once get

$$r q_r = (-1)^{r+1} \frac{\sum (-1)^{r+k}(k-1)! r Q_A^{k_1} Q_B^{k_2} \cdots}{k_1! k_2! \cdots}$$

which is the result (A) obtained by O'Toole.

From (C), D_r follows in terms of d_r and thence with (2) Q_r can be expressed in terms of q_r. Using multinomial theorem we arrive at

$$r! Q_r = \frac{\sum r! d_A^{k_1} d_B^{k_2} \cdots}{k_1! k_2! \cdots}$$

which is (B). Hence both the results of O'Toole have been deduced.

3. In his second paper[5] O'Toole defines symmetric functions for more than one system of Variates. I call such symmetric functions *Hyper Symmetric Functions.*

The Hyper operators are developed to express Hyper symmetric functions in terms of hyper power-sums. They are defined by O'Toole by the following relations, taking into consideration two systems of variates only,

$$d_{pq}^r = \frac{d^r}{ds_{pq}};$$

---

[4] p. 29.

[5] Ann. Stat. 3. (1932), 56-63.

(A)' $$d_{pq} = \frac{\sum (-1)^{k+1}(k-1)!\,D^{k_1}_{p_1q_1} D^{k_2}_{p_2q_2} \cdots}{k_1!\,k_2! \cdots}$$

(B') $$D_{pq} = \frac{\sum d^{k_1}_{p_1q_1} d^{k_2}_{p_2q_2} \cdots}{k_1!\,k_2! \cdots}$$

$$\text{where } k_1p_1 + k_2p_2 + \cdots = p$$

$$k_1q_1 + k_2q_2 + \cdots = q$$

These relations readily follow from Macmahon's[6] hyper operators g_{pq} and G_{pq}. These operators came into existence with the problem of expressing hyper symmetric functions in terms of hyper elementary symmetric functions and they are connected by the following relations.

I. $$(-1)^{p+q-1} \frac{(p+q-1)!}{p!\,q!} g_{pq} = \frac{\sum (-1)^{k-1}(k-1)!}{k_1!\,k_2! \cdots} G^{k_1}_{p_1q_1} G^{k_2}_{p_2q_2} \cdots$$

II. $$(-1)^{p+q-1} G_{pq} = \sum \frac{[(p_1+q_1-1)!]^{k_1}}{p_1!\,q_1!} \frac{[(p_2+q_2-1)!]^{k_2}}{p_2!\,q_2!} \cdots \frac{(-1)^{k-1}}{k_1!\,k_2! \cdots} (g_{p_1q_1})^{k_1} (g_{p_2q_2})^{k_2} \cdots.$$

Macmahon[7] has also shown that

$$g_{pq} s_{pq} = (-1)^{p+q-1} \frac{p!\,q!}{(p+q-1)!};$$

from which we get

$$g_{pq} = (-1)^{p+q-1} \frac{p!\,q!}{(p+q-1)!} d_{pq} \tag{3}$$

The operator G behaves like D of O'Toole. Now using (3) we derive from (I) the result (A') arrived at by O'Toole without reference to Macmahon. Similarly from II, using (3) (B') is deduced.

UNIVERSITY COLLEGE, SWANSEA, WALES.

---

[6] Macmahon, Comb. Analysis. Vol. II. Cambridge University Press (1916), p. 302.

[7] Macmahon Op. Cit., p. 304.

GRADUATION BY A TRUNCATED NORMAL

By Nathan Keyfitz

Below is a table for finding the constants of a truncated normal by the equation of moments. Karl Pearson* gives such a table for the case in which the data are to be fitted to the "tail" (i.e. less than half) of a normal curve but I do not believe that the formulae for a distribution consisting of more than half of a normal curve have before been tabulated.

The table below was calculated primarily for an investigation being carried out on the duration of unemployment. The Canadian Census of 1931 reported the number of persons losing 1–4, 5–8, … 49–52 weeks in the course of the year June 1st, 1930 to June 1st, 1931, by various classifications, (industry, province, age, etc.).

The tendency to report even numbers of months on the part of the enumerated population was evident in the result, and some kind of graduation was necessary for an interpretation. After some experiment a part of a normal curve was settled upon as the simplest and generally most satisfactory representation.

It was found that among the classes of workers in which unemployment is high the curve is more advanced,—i.e. the mode is at a higher number of weeks,—than in the classes where unemployment is low. In many cases, (in most groupings of female workers for instance) where unemployment is relatively very low the modal point of the uncurtailed normal stands at a negative number of weeks,—for these cases the fitting is to a true tail and the tables of the Biometric Laboratory were used.

Details of the results of the investigation will be published shortly in the Unemployment Monograph of the Dominion Bureau of Statistics. Meanwhile, this table will be of use as the complement of Pearson's tabulation which is only suitable for $\psi_1 \geqq .5708$.

Table for finding the constants of a truncated normal by the equation of moments

x'	ψ_1	$\Delta\psi_1$	ψ_2	$\Delta\psi_2$
0	.5708	−.0180	1.2533	−.0562
.1	.5528	−.0183	1.1971	−.0543
.2	.5345	−.0188	1.1428	−.0526
.3	.5157	−.0190	1.0902	−.0506
.4	.4967	−.0193	1.0396	−.0487
.5	.4774	−.0195	.9909	−.0467

* Tables for Statisticians and Biometricians, page 25.

χ'	ψ_1	$\Delta\psi_1$	ψ_2	$\Delta\psi_2$
.6	.4579	−.0195	.9442	−.0449
.7	.4384	−.0196	.8993	−.0428
.8	.4188	−.0196	.8565	−.0409
.9	.3992	−.0194	.8156	−.0390
1.0	.3798	−.0192	.7766	−.0370
1.1	.3606	−.0189	.7396	−.0351
1.2	.3417	−.0185	.7045	−.0332
1.3	.3232	−.0180	.6713	−.0315
1.4	.3052	−.0175	.6398	−.0296
1.5	.2877	−.0170	.6102	−.0279
1.6	.2707	−.0163	.5823	−.0263
1.7	.2544	−.0156	.5560	−.0246
1.8	.2388	−.0148	.5314	−.0232
1.9	.2240	−.0141	.5082	−.0216
2.0	.2099	−.0134	.4866	−.0204
2.1	.1965	−.0126	.4662	−.0190
2.2	.1839	−.0118	.4472	−.0178
2.3	.1721	−.0110	.4294	−.0166
2.4	.1611	−.0103	.4128	−.0156
2.5	.1508	−.0096	.3972	−.0146
2.6	.1412	−.0089	.3826	−.0137
2.7	.1323	−.0083	.3689	−.0128
2.8	.1240	−.0076	.3561	−.0120
2.9	.1164	−.0071	.3441	−.0113
3.0	.1093		.3328	
3.5	.0813		.2856	
4.0	.06246		.24999	
4.5	.049379		.222221	
5.0	.0399997		.1999999	

Let d = distance of centroid of actual distribution from point of truncation.

Let Σ = standard deviation of distribution about its mean. Then $\psi_1 = \dfrac{\Sigma^2}{d^2}$.

Hence corresponding χ' and ψ_2 may be found.
Then $\sigma = d\psi_2$, *where* σ = *standard deviation of uncurtailed normal.*
And $\chi = \chi'\sigma$, where χ = origin of uncurtailed normal.
N.B. The point of truncation is taken for the origin in the original distribution.

Social Analysis Branch, Dominion Bureau of Statistics,
Ottawa, Canada

REPORT OF THE ANNUAL MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The annual meeting of the Institute of Mathematical Statistics was held on Wednesday and Thursday, December 29-30, 1937, in Indianapolis, Indiana, in conjunction with the meetings of the American Mathematical Society and associated organizations.

The Wednesday morning session was devoted to applications of statistics to industry and engineering. On Thursday morning, the Institute held a joint session with the Mathematical Society for the presentation of voluntary papers on probability and statistics. This session was immediately followed by another of the Institute for two invited addresses. These addresses were "The theory of general means" by Professor E. L. Dodd, and "On the independence of certain estimates of variance" by Professor A. T. Craig. Professor P. R. Rider was in charge of arranging the program.

On Thursday noon, there was a luncheon at the Marott Hotel for members of the Institute and their guests. After the luncheon, Professor H. L. Rietz spoke on "The future of the Institute in relation to mathematical statistics."

At the business meeting, which followed the Wednesday morning session, President Shewhart announced that these officers had been elected for 1938: President, B. H. Camp, Wesleyan University; Vice-Presidents, P. R. Rider, Washington University, and S. S. Wilks, Princeton University; Secretary-Treasurer, A. T. Craig, University of Iowa. The Institute voted to hold its 1938 meeting with the American Statistical Association. The meeting will be in Detroit, Michigan, in December of this year.

ALLEN T. CRAIG, *Secretary.*

THE ANNALS *of* MATHEMATICAL STATISTICS

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. IX, No. 2 — June, 1938

TESTS OF STATISTICAL HYPOTHESES WHICH ARE UNBIASED IN THE LIMIT

By J. Neyman

1. Introduction. The idea of unbiased tests of statistical hypotheses has been put forward and discussed in two recent papers.[1] Recently also a particular problem was solved introducing a test which has the property of being unbiased in the limit.[2] The purpose of the present note is to discuss this conception in its general form and to indicate methods of determining the tests unbiased in the limit of a broad class of simple statistical hypotheses. The notation and the terminology employed below are explained in the papers quoted.

2. Notation and definitions. Consider a set of n random variables

$$X_1, X_2, \cdots X_n \tag{1}$$

the particular values of which

$$x_1, x_2, \cdots x_n \tag{2}$$

can be given by observation and denote by Ω the set of hypotheses concerning the probability law of (1) which are regarded as admissible. We shall assume that all the hypotheses included in Ω specify the probability law of the X's having the same analytical form but differing among them in the value of just one parameter, θ. Thus, if E_n denotes the point (the "event point") in the space W_n of n dimensions with its coordinates equal to the values of (1) and w_n any region in W_n, then the probability of E_n falling within w_n, as determined by any of the hypotheses forming the set Ω will be denoted by

$$P\{E_n \epsilon w_n \mid \theta\} \tag{3}$$

and will be a function of the parameter θ. The probability (3) with fixed θ considered as a function of varying w_n is called the integral probability law of the X's. Frequently (3) is equal to the integral of a certain non-negative function of E_n over the region w_n. This function will always be denoted by $p(E_n \mid \theta)$ and called the elementary probability law of (1).

---

[1] J. Neyman and E. S. Pearson: Contributions to the Theory of Testing Statistical Hypotheses. Part I. Stat. Res. Memoirs, Vol. 1, (1936) pp. 1–37. Part II, ibid., Vol. II (1938).

J. Neyman: Sur la vérification des hypothèses statistiques composées. Bull. Soc. Math. de France, Vol. 63 (1935), pp. 246–266.

[2] J. Neyman: "Smooth" Test for Goodness of Fit. Skandinavisk Aktuarietidskrift, (1937), pp. 149–199.

Denote by H_0 some particular hypothesis of the set Ω and by θ_0 the value that it ascribes to the parameter θ.

A test of the statistical hypothesis H_0 consists in a rule of rejecting H_0 whenever E_n falls within a specified region w_n and in not doing so in other cases. The region w_n used for this purpose is called the critical region. It follows that to choose a test means to choose a critical region.

We shall consider below only cases such that for any region w_n the probability (3) considered as a function of θ possesses two successive derivatives.

DEFINITION 1. *If a critical region $\bar{w}_n$ has the property that, α being a fixed positive number,*

$$(4) \qquad \text{(a)} \quad P\{E_n \epsilon \bar{w}_n \mid \theta_0\} = \alpha$$

$$(5) \qquad \text{(b)} \quad \frac{d}{d\theta} P\{E_n \epsilon \bar{w}_n \mid \theta\}\Big|_{\theta=\theta_0} = 0$$

$$(6) \qquad \text{(c)} \quad \frac{d^2}{d\theta^2} P\{E_n \epsilon \bar{w}_n \mid \theta\}\Big|_{\theta=\theta_0} \geq \frac{d^2}{d\theta^2} P\{E_n \epsilon w_n \mid \theta\}\Big|_{\theta=\theta_0}$$

where w_n is any region satisfying (a) *and* (b), *then the region $\bar{w}_n$ is called the unbiased critical region of type A corresponding to the level of significance α, and the test of the hypothesis H_0 based on $\bar{w}_n$, the unbiased test of type A.*

This is the definition given in the first of the earlier papers quoted. Now we shall define the test which is unbiased in the limit. For this purpose we shall have to consider the situation where n is indefinitely increased and consequently we have a sequence of probability laws (3), a sequence of spaces W_n where they are defined and a sequence of regions $\bar{w}_n$, each $\bar{w}_n$ being a part of the corresponding W_n.

We must also introduce a varying scale with which to measure the differences $\theta - \theta_0$. This is due to the fact that, if the choice of the sequence of regions $\bar{w}_n$ is not very unlucky and $\theta \neq \theta_0$, then we shall frequently have

$$(7) \qquad \lim_{n\to\infty} P\{E_n \epsilon \bar{w}_n \mid \theta\} = 1$$

Comparing this with condition (4), we see that in general the limit of

$$P\{E_n \epsilon \bar{w}_n \mid \theta\}$$

for $n \to \infty$ will be discontinuous at $\theta = \theta_0$. To avoid this we shall measure $\theta - \theta_0$ in terms of $n^{-\frac{1}{2}}$ introducing instead of θ a new parameter ϑ connected with the former by means of the equality

$$(8) \qquad \theta = \theta_0 + \frac{\vartheta}{\sqrt{n}}$$

For the hypothesis tested H_0 we shall have $\vartheta = 0$ and $\vartheta \neq 0$ for any other hypothesis in Ω. The new parameter ϑ thus introduced will be called the

standardized error in H_0. It will be frequently convenient to use θ but occasionally we shall use ϑ as well, for example writing $P\{E_n \epsilon w_n \mid \vartheta\}$ instead of (3) etc., and it is necessary to remember the connection (8) existing between θ and ϑ. It may be useful to notice at once that $df/d\theta = \sqrt{n}\, df/d\vartheta$.

DEFINITION 2. *We shall say that the sequence of regions*

$$\bar{w}_1, \bar{w}_2, \cdots, \bar{w}_n, \cdots \tag{9}$$

determines a test of the hypothesis H_0 which is unbiased in the limit and corresponds (in the limit) to the level of significance α, if for any n

$$\text{(d)} \quad \frac{d^2}{d\vartheta^2} P\{E_n \epsilon \bar{w}_n \mid \vartheta\}\Big|_{\vartheta=0} \geq \frac{d^2}{d\vartheta^2} P\{E_n \epsilon w_n \mid \vartheta\}\Big|_{\vartheta=0} \tag{10}$$

where w_n is any region such that

$$P\{E_n \epsilon w_n \mid \vartheta = 0\} = P\{E_n \epsilon \bar{w}_n \mid \vartheta = 0\} \tag{11}$$

and

$$\frac{d}{d\vartheta} P\{E_n \epsilon w_n \mid \vartheta\}\Big|_{\vartheta=0} = \frac{d}{d\vartheta} P\{E_n \epsilon \bar{w}_n \mid \vartheta\}\Big|_{\vartheta=0} \tag{12}$$

and if

$$\text{(e)} \quad \lim_{n \to \infty} P\{E_n \epsilon \bar{w}_n \mid \vartheta = 0\} = \alpha \tag{13}$$

$$\text{(f)} \quad \lim_{n \to \infty} \frac{d}{d\vartheta} P\{E_n \epsilon \bar{w}_n \mid \vartheta\}_{\vartheta=0} = 0 \tag{14}$$

The practical application of the test determined by the sequence of regions (7) consists in observing as large a number n of the X's of (1) and in rejecting the hypothesis H_0 whenever E_n falls within $\bar{w}_n$. If n is sufficiently large, then this rule will have about the same advantages as the application of the unbiased test of type A. In fact, allowing for the circumstance that the values of (11) and (12) will be only approximately equal to the limits (13) and (14), the properties of the test satisfying the Definition 2 will be as follows: If the hypothesis tested be true, it will be wrongly rejected with a relative frequency approximately equal to α fixed in advance. If H_0 is false and the true value say ϑ' of ϑ is not very different from zero, then the frequency of rejecting H_0 will be greater than α and could not be increased by applying some other similar test.

It may be useful to notice that in general there may be more than one test of the same hypothesis which is unbiased in the limit and corresponds to a fixed level of significance. Consequently there is a possibility of choosing between such tests, but it seems to the author that such a choice would require a previous strengthening of the theorem of S. Bernstein on which the present work is based.

3. Theorem of S. Bernstein. In the following, we shall have to use the following particular case of a theorem due to S. Bernstein.[3] Denote by $\mathcal{E}(x)$ the mathematical expectation of any variate x and by

$$\begin{matrix} X_1, X_2, \cdots, X_n, \cdots \\ Y_1, Y_2, \cdots, Y_n, \cdots \end{matrix} \tag{15}$$

two unlimited sequences of random variables.

We shall assume that

(1) X_i is independent of X_j and Y_j for any $i \neq j$.

(2) The following mathematical expectations exist and are independent of i:

$$\begin{aligned} &\mathcal{E}(X_i) = a \qquad \mathcal{E}(Y_i) = b \\ &\mathcal{E}(X_i - a)^2 = \sigma_1^2 \qquad \mathcal{E}(Y_i - b)^2 = \sigma_2^2 \\ &\mathcal{E}[(X_i - a)(Y_i - b)] = r\sigma_1\sigma_2 \\ &\mathcal{E}(| X_i - a |^3) = \mu \quad \mathcal{E}(| Y_i - b |^3) = \nu \end{aligned} \tag{16}$$

Consider now the space of $2n$ dimensions W_n and denote by E_n a point in it as determined by the values of X_i, Y_i for $i = 1, 2, \cdots n$ considered as its coordinates. Let u_n and v_n denote the sums

$$u_n = \sum_{i=1}^{n} X_i, \qquad v_n = \sum_{i=1}^{n} Y_i \tag{17}$$

and denote by D_n the point on a plane S with its orthogonal coordinates equal to u_n and v_n. If s is any region in S then let $P\{D_n \epsilon s\}$ be the probability of D_n falling within s.

Theorem of S. Bernstein. *If the variates* (15) *satisfy the conditions* (1) *and* (2) *then, for any* $\epsilon > 0$, *there exists a number* N_ϵ, *such that the inequality* $n > N_\epsilon$ *implies*

$$\left| P\{D_n \epsilon s\} - \frac{1}{2\pi n\sigma_1\sigma_2\sqrt{1-r^2}} \iint_s e^{-\frac{1}{2n(1-r^2)}\left(\frac{(u-na)^2}{\sigma_1^2} - 2r\frac{u-na}{\sigma_1}\frac{v-nb}{\sigma_2} + \frac{(v-nb)^2}{\sigma_2^2}\right)} du\, dv \right| < \epsilon, \tag{18}$$

whatever the region s *in* S *may be.*

4. Tests unbiased in the limit. We shall consider the problem of determining the tests satisfying Definition 2, in the case where the following hypotheses are fulfilled.

---

[3] S. Bernstein: Sur un théorème limite du calcul des probabilités. Math. Ann., Bd. 97 (1926) p. 44.

See also V. Romanovskij, Bull. de l'Académie des Sciences de l'U. R. S. S., 1929, p. 209 and W. Kozakiewicz, Ann. Soc. Polonaise Math., t. XIII (1934), pp. 24–43.

(i) All the random variables (1) are mutually independent and each of them follows the same elementary probability law which we shall denote by $p(x_i \mid \theta)$.

(ii) The elementary probability law $p(x_i \mid \theta)$ admits three differentiations and two consecutive differentiations with respect to θ under the integral taken over any fixed finite or infinite interval, so that

$$\frac{d^k}{d\theta^k} \int_a^b p(x_i \mid \theta)\, dx_i = \int_a^b \frac{d^k}{d\theta^k} p(x_i \mid \theta)\, dx_i \tag{19}$$

for $k = 1, 2$.

(iii) If

$$\varphi_i = \left.\frac{\partial \log p(x_i \mid \theta)}{\partial \theta}\right|_{\theta=\theta_0} \quad \text{and} \quad \Psi_i = \left.\frac{\partial^2 \log p(x_i \mid \theta)}{\partial \theta^2}\right|_{\theta=\theta_0}, \tag{20}$$

then we shall assume the existence of the following integrals all taken from $-\infty$ to $+\infty$

$$\sigma_1^2 = \int \varphi_i^2 p(x_i \mid \theta_0)\, dx_i \tag{21}$$

$$\sigma_2^2 = \int (\Psi_i + \sigma_1^2)^2\, p(x_i \mid \theta_0)\, dx_i \tag{22}$$

$$r\sigma_1\sigma_2 = \int \varphi_i \Psi_i p(x_i \mid \theta_0)\, dx \tag{23}$$

$$\mu = \int \mid \varphi_i^3 \mid p(x_i \mid \theta_0)\, dx_i \tag{24}$$

$$\nu = \int \mid \Psi_i + \sigma_1^2 \mid^3 p(x_i \mid \theta_0)\, dx_i \tag{25}$$

PROPOSITION I. *If the above conditions* (i), (ii) *and* (iii) *are satisfied*, Ψ_i *being a function of* x_i *and* $\mid r \mid < 1$, *then the sequence of regions* ϖ_n *including all the points of* W_n *where* $p(E_n \mid \theta_0) = 0$ *and also those of the remaining ones which satisfy the inequality*

$$\sum_{i=1}^{n} \Psi_i + \left(\sum_{i=1}^{n} \varphi_i\right)^2 \geq M\sigma_2\sqrt{n(1-r^2)} - n\sigma_1^2 + r\frac{\sigma_2}{\sigma_1}\sum_{i=1}^{n} \varphi_i \tag{26}$$

where the coefficient M is to be found from the equation

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}x^2} \frac{1}{\sqrt{2\pi}} \int_{M-Nx^2}^{\infty} e^{-\frac{1}{2}y^2}\, dy \right\} dx = \alpha \tag{27}$$

with

$$N = \frac{\sigma_1^2\sqrt{n}}{\sigma_2\sqrt{1-r^2}}, \tag{28}$$

defines a test of the hypothesis H_0, which is unbiased in the limit and corresponds (in the limit) to the level of significance α.

REMARK. The calculation of M satisfying the equation (27) is, of course, laborious. But a table of values of M corresponding to varying values of N is being constructed by N. L. Johnson at the Department of Statistics, University College, London, and it is hoped that it will soon be published.

To prove Proposition I, we must first prove (a) that whatever n, the region $\bar{w}_n$ determined by the inequality (26) satisfies the condition (d) in the definition 2. The proof is based on the following Lemma.[4]

LEMMA. *If $F_0, F_1, \cdots F_m$ are functions of $x_1, \cdots x_n$ integrable over any region in W_n and w_0 a region in W_n such that within w_0*

$$F_0 \geq \sum_{i=1}^{m} a_i F_i \tag{29}$$

while outside of w_0

$$F_0 \leq \sum_{i=1}^{m} a_i F_i \tag{30}$$

$a_1, a_2 \cdots a_m$ being some constant coefficients, then, whatever may be any other region w in W_n, such that

$$\int \cdots \int_w F_i dx_1 \cdots dx_n = \int \cdots \int_{w_0} F_i dx_1 \cdots dx_n, \quad \text{for } i = 1, 2, \cdots m, \tag{31}$$

we shall have

$$\int \cdots \int_{w_0} F_0 \, dx_1 \cdots dx_n \geq \int \cdots \int_w F_0 \, dx_1 \cdots dx_n. \tag{32}$$

PROOF OF PROPOSITION I. Denote, for simplicity, by $p(E_n)$ the elementary probability law of the X's as determined by the hypothesis tested. Comparing the statement of the Lemma with the definition (26) of $\bar{w}_n$, we immediately see that this region has the following property: whatever may be any other region w in W_n such that

$$\int \int_w p(E_n) \, dx_1 \cdots dx_n = \int \cdots \int_{\bar{w}_n} p(E_n) \, dx_1 \cdots dx_n \tag{33}$$

and

$$\frac{1}{\sqrt{n}} \int \cdots \int_w \sum_{i=1}^{n} \varphi_i p(E_n) \, dx_1 \cdots dx_n = \frac{1}{\sqrt{n}} \int \cdots \int_{\bar{w}_n} \sum_{i=1}^{n} \varphi_i p(E_n) \, dx_1 \cdots dx_n \tag{34}$$

[4] J. Neyman and E. S. Pearson: loc. cit., pp. 10–11.

we shall have

$$\frac{1}{n}\int \cdots \int_{\bar{w}_n} \left(\sum_{i=1}^{n} \Psi_i + \left(\sum_{i=1}^{n} \varphi_i\right)^2\right) p(E_n)\, dx_1, \cdots dx_n \geq \frac{1}{n}\int \cdots \int_{w} \left(\sum_{i=1}^{n} \Psi_i + \left(\sum_{i=1}^{n} \varphi_i\right)^2\right) p(E_n)\, dx_1 \cdots dx_n \tag{35}$$

But under the conditions (i) and (ii)

$$p(E_n \mid \vartheta) = \prod_{i=1}^{n} p(x_i \mid \theta_0 + \vartheta/\sqrt{n}) \tag{36}$$

$$\left.\frac{\partial p(E_n \mid \vartheta)}{\partial \vartheta}\right|_{\vartheta=0} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_i p(E_n) \tag{37}$$

$$\left.\frac{\partial^2 p(E_n \mid \vartheta)}{\partial \vartheta^2}\right|_{\vartheta=0} = \frac{1}{n}\left(\sum_{i=1}^{n} \Psi_i + \left(\sum_{i=1}^{n} \varphi_i\right)^2\right) p(E_n) \tag{38}$$

and it is easily seen that the relations (33), (34) and (35) are identical with (11), (12) and (10) respectively and that therefore the region $\bar{w}_n$ satisfies the condition (d) of definition 2. It remains to prove that $\bar{w}_n$ satisfies also the conditions (e) and (f), that is to say that, for $n \to \infty$, the formulas in the right hand sides of (33) and (34) tend to the prescribed values α and zero respectively. This conclusion concerning (33) is a consequence of the theorem of S. Bernstein, quoted above. To see this, write

$$u_n = \sum_{i=1}^{n} \varphi_i, \qquad v_n = \sum_{i=1}^{n} \Psi_i \tag{39}$$

and denote by s_0 the region in the plane S of (u, v) defined by the inequality

$$v + u^2 \geq M\sigma_2\sqrt{n(1 - r^2)} - n\sigma_1^2 + r\frac{\sigma_2}{\sigma_1}u \tag{40}$$

obtained from (26) by means of (39). The right hand side of (33) represents the probability determined by the hypothesis tested of the X's satisfying the inequality (26). But this is satisfied simultaneously with the variates u_n and v_n satisfying (40). Therefore, if we denote by D_n the point in S with its coordinates equal to (39), then the right hand side of (33) may be interpreted as the probability $P\{D_n \epsilon s_0\}$ of D_n falling within s_0. Comparing (21)–(25) with (16), it is easily seen that, according to the Theorem of S. Bernstein, whatever may be $\epsilon > 0$, if n is sufficiently large, then

$$\left| P\{D_n \epsilon s_0\} - \iint_{s_0} G_n\, dv \right| < \epsilon \tag{41}$$

where

$$G_n = \frac{1}{2\pi n\sigma_1\sigma_2\sqrt{1 - r^2}}\, e^{-\frac{1}{2n(1-r^2)}\left\{\frac{u^2}{\sigma_1^2} - 2r\frac{u}{\sigma_1}\,\frac{v+n\sigma_1^2}{\sigma_2} + \frac{(v+n\sigma_1^2)^2}{\sigma_2^2}\right\}} \tag{42}$$

In fact, to what is given explicitly, we must only add that as

$$\int_{-\infty}^{+\infty} p(x_i \mid \theta)\, dx_i = 1 \tag{43}$$

the derivative with respect to θ of the left hand side must be identically equal to zero. Therefore

$$\frac{d}{d\theta} \int p(x_i \mid \theta)\, dx_i \Big|_{\theta=\theta_0} = \int \varphi_i p(x_i \mid \theta_0)\, dx_i = \mathcal{E}(\varphi_i) = 0 \tag{44}$$

where again the integrals are taken from $-\infty$ to $+\infty$. It follows further that the second derivative with respect to θ of (43) must be again identically equal to zero. Therefore, keeping in mind the definitions of φ_i and Ψ_i, we may write

$$\frac{\partial^2}{\partial\theta^2} \int p(x_i \mid \theta)\, dx_i \Big|_{\theta=\theta_0} = \int (\Psi_i + \varphi_i^2) p(x_i \mid \theta_0)\, dx_i = 0 \tag{45}$$

and thus

$$\mathcal{E}(\Psi_i) = -\mathcal{E}(\varphi_i^2) = -\sigma_1^2 \tag{46}$$

The proof that the right hand side of (24) tends to α with $n \to \infty$ will be completed if we manage to reduce the integral of (42) over the region s_0 to the integral (27). This is easily done by substituting

$$x = \frac{u}{\sigma_1\sqrt{n}} \tag{47}$$

$$y = \frac{v + n\sigma_1^2 - r\sigma_2 u/\sigma_1}{\sigma_2\sqrt{n(1 - r^2)}} \tag{48}$$

Thus, if the coefficient M in (26) and (40) satisfies the condition (27), then the value of the integral of G_n in (41) is permanently equal to α and this means that the right hand side of (33) tends to α as $n \to \infty$.

Denote by $p_n(u, v)$ the elementary probability law of u_n and v_n. It will be noticed that, whatever s in S

$$P\{D_n \in s\} = \iint_s p_n(u, v)\, du\, dv \tag{49}$$

and that consequently in the course of the above discussion we have proved that, whatever $\epsilon > 0$, there exists a sufficiently large number N_ϵ such that $n > N_\epsilon$ implies

$$\left| \iint_s (p_n(u, v) - G_n)\, du\, dv \right| < \epsilon \tag{50}$$

whatever may be the region s in S. We shall now use this circumstance to prove that, when $n \to \infty$, the right hand side of (34) tends to zero. It will be noticed first that

$$(51) \qquad \int \cdots \int_{\bar{w}_n} (\sum \varphi_i)^k p(E_n)\, dx_1 \cdots dx_n = \int\int_{s_0} u^k p_n(u, v)\, du\, dv$$

for $k = 1, 2$. Further

$$(52) \qquad \int\int_{s_0} u^2 p_n(u, v)\, du\, dv \leq \int\int_S u^2 p_n(u, v)\, du\, dv = n\sigma_1^2$$

Using the inequality of Schwartz,[5] we may write

$$(53) \qquad \begin{aligned} &\left| \frac{1}{\sqrt{n}} \int\int_{s_0} u(p_n(u, v) - G_n)\, du\, dx \right| \\ &\quad \leq \frac{1}{\sqrt{n}} \left(\int\int_{s_0} u^2 \,|\, p_n(u, v) - G_n \,|\, du\, dv \int\int_{s_0} |\, p_n(u, v) - G_n \,|\, du\, dv \right)^{\frac{1}{2}} \end{aligned}$$

Now, it is easy to calculate that

$$(54) \qquad \int\int_{s_0} u^2 \,|\, p_n(u, v) - G_n \,|\, du\, dv \leq 2n\sigma_1^2$$

On the other hand, if n is so large that (50) holds good for any region s in S and s_+ and s_- denote the two parts of s_0 where $p_n(u, v) - G_n$ is respectively positive and negative, then

$$(55) \qquad \begin{aligned} 0 \leq \int\int_{s_0} |\, p_n(u, v) - G_n \,|\, du\, dv &= \int\int_{s_+} (p_n(u, v) - G_n)\, du\, dv \\ &\quad - \int\int_{s_-} (p_n(u, v) - G_n)\, du\, dv < 2\epsilon \end{aligned}$$

and it follows that, for such large values of n,

$$(56) \qquad \left| \frac{1}{\sqrt{n}} \int\int_{s_0} u(p_n(u, v) - G_n)\, du\, dv \right| \leq 2\sigma_1 \sqrt{\epsilon}$$

On the other hand, using the transformation (47) and (48), we find that

$$(57) \qquad \frac{1}{\sigma_1 \sqrt{n}} \int\int_{s_0} u G_n\, du\, dv = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ x e^{-\frac{1}{2}x^2} \frac{1}{\sqrt{2\pi}} \int_{M-Nx^2}^{\infty} e^{-\frac{1}{2}y^2}\, dy \right\} dx$$

and consequently is permanently equal to zero. As ϵ is an arbitrarily small number, it follows that

$$(58) \qquad \lim_{n\to\infty} \frac{1}{\sqrt{n}} \int\int_{s_0} u p_n(u, v)\, du\, dv = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \int\int_{\bar{w}_n} \sum_{i=1}^{n} \varphi_i p(E_n)\, dx_1 \cdots dx_n = 0$$

which fulfills the proof of Proposition I.

---

[5] See for example: S. Kaczmarz and H. Steinhaus, Theorie der Orthogonalreihen, Warsaw, 1935, p. 10.

Proposition II. *If the conditions of Proposition I are satisfied but either $|r| = 1$ or Ψ_i is independent of x_i, then the test of the hypothesis H_0 which is unbiased in the limit and which corresponds, to the level of significance α, is determined by the sequence of critical regions $\bar{w}_n$, defined by the inequality*

$$\left|\sum_{i=1}^{n} \varphi_i\right| \geq \lambda\sigma_1\sqrt{n} \tag{59}$$

where λ satisfies the equation

$$\frac{1}{\sqrt{2\pi}}\int_{-\lambda}^{+\lambda} e^{-\frac{1}{2}x^2}\,dx = 1 - \alpha \tag{60}$$

Proof. We notice first that the condition $|r| = 1$ and the equation (44) imply

$$\left(\int \varphi_i(\Psi_i + \sigma_1^2)p(x_i \mid \theta_0)\,dx_i\right)^2 = \int \varphi_i^2 p(x_i \mid \theta_0)\,dx_i \int (\Psi_i + \sigma_1^2)^2 p(x_i \mid \theta_0)\,dx_i \neq 0 \tag{61}$$

or

$$\frac{\int \varphi_i(\Psi_i + \sigma_1^2)p(x_i \mid \theta_0)\,dx_i}{\int \varphi_i^2 p(x_i \mid \theta_0)\,dx_i} = \frac{\int (\Psi_i + \sigma_1^2)^2 p(x_i \mid \theta_0)\,dx_i}{\int \varphi_i(\Psi_i + \sigma_1^2)p(x_i \mid \theta_0)\,dx_i} = A \neq 0 \tag{62}$$

and therefore

$$\int \{(\Psi_i + \sigma_1^2)^2 - A\varphi_i(\Psi_i + \sigma_1^2)\}p(x_i \mid \theta_0)\,dx_i = 0 \tag{63}$$

$$\int \{\varphi_i(\Psi_i + \sigma_1^2) - A\varphi_i^2\}p(x_i \mid \theta_0)\,dx_i = 0 \tag{64}$$

and finally

$$\int (\Psi + \sigma_1^2 - A\varphi_i)^2 p(x_i \mid \theta_0)\,dx_i = 0 \tag{65}$$

which means that at almost every value of x_i for which $p(x_i \mid \theta_0) \neq 0$,

$$\Psi_i + \sigma_1^2 = A\varphi_i \tag{66}$$

It follows that the inequality (10) in the definition 2 of the test which is unbiased in the limit reduces to the following

$$\frac{1}{n}\int \cdots \int_{\bar{w}_n} \left(\sum \varphi_i\right)^2 p(E_n \mid \theta_0)\,dx_1 \cdots dx_n \geq \frac{1}{n}\int \cdots \int_{w} \left(\sum \varphi_i\right)^2 p(E_n \mid \theta_0)\,dx_1 \cdots dx_n \tag{67}$$

owing to (11), (12), (37) and (38). On the other hand, the inequality (59) is equivalent to

$$(\sum \varphi_i)^2 \geq a + b \sum \varphi_i \tag{68}$$

with $a = \lambda^2\sigma_1^2 n$ and $b = 0$. Referring to the Lemma, we conclude that the regions $\bar{w}_n$ satisfy the condition (d) of the Definition 2. It remains to show that they satisfy also the conditions (e) and (f). This immediately follows from the theorem of Liapounoff[6] and the reasoning which we used above in order to prove (58).

If Ψ_i does not depend on x_i then, owing to (38) and (11), the inequality (10) immediately reduces to (67) and the proof of Proposition II follows exactly the same lines as before.

5. **Limiting power function.** To know the properties of a test undoubtedly means to know (i) how frequently this particular test will reject the hypothesis tested when it is in fact true and (ii) how frequently will it detect its falsehood when it is wrong. The information of this kind is provided by the properties of the so called power function of the test. This has been defined[7] as follows. Let w_n be any critical region and, as formerly, $P\{E_n \epsilon w_n \mid \theta\}$ the probability of E_n falling within w_n as determined by a specified value of θ. If w_n is fixed, then $P\{E_n \epsilon w_n \mid \theta\}$ will be a function of θ only. To emphasize this circumstance we may introduce a new symbol, writing

$$P\{E_n \epsilon w_n \mid \theta\} = \beta(\theta \mid w_n\} \tag{69}$$

which will mean that in the above formula w_n is kept constant and θ varied. The function $\beta(\theta \mid w_n)$ thus defined is called the power function of the critical region w_n or that of the test based on w_n. If w_n corresponds to the level of significance α and θ_0 is the value of θ specified by the hypothesis tested H_0, then

$$\beta(\theta_0 \mid w_n) = \alpha \tag{70}$$

and it will be noticed that this is the probability of rejecting H_0 when it is in fact true. As we reject H_0 only in such cases when $E_n \epsilon w_n$, the values of $\beta(\theta \mid w_n)$ corresponding to other values of $\theta \neq \theta_0$ are equal to the probability of detecting the falsehood of the hypothesis H_0 when θ has any specified value different from θ_0. The larger the value of $\beta(\theta \mid w_n)$ at a given θ, the greater will be the "detecting power" of the test, which justifies the name attached to the function $\beta(\theta \mid w_n)$. Until the present time the power function of only a few tests has been studied and it follows that we know comparatively little of the properties of the tests even if they are in frequent use. The first study of this kind was concerned with the power function of the "Student's" test as applied to the problem of one sample and there are three publications giving various

[6] See for example Paul Lévy: Théorie de l'addition des variables aléatoires. Paris, 1937. Pp. 101–107.

[7] J. Neyman and E. S. Pearson: loc. cit., p. 9.

numerical tables.[8] However, in these publications the term "power function" does not appear yet. Apart from the joint paper already referred to where the term "power function" was first defined, we may mention a few papers in *Biometrika*, the most important of which seems to be that by S. S. Wilks and Catherine M. Thompson.[9] The purpose of studying the power function of any test is to be able to answer the following three questions:

(a) What should be the size of a sample in order to have a reasonable chance of detecting the falsehood of the hypothesis tested, when the error in the parameters that it specifies has some stated value?

(b) If in some particular case a test failed to reject the hypothesis tested (which, of course, does not mean that it is necessarily true), is it likely that the error in θ_0 does not exceed some specified limit Δ?

(c) Two different tests corresponding to the same level of significance are suggested for the same hypothesis H_0, which shall we use?

In this last case the answer is obvious—the one which gives the greater chance of detecting the falsehood of the hypothesis tested in cases when it is wrong. But to know this we must know the power functions of both tests.

For the above reasons it seems to be important to study the power function of the test unbiased in the limit as defined above. It is obvious that, as in this case the elementary probability laws are not specified, it is impossible to find the actual explicit formula giving the power function. Therefore we shall endeavour to find its limiting form. This will be done by means of the two following theorems.

Consider an infinite sequence of situations

$$S_1, S_2, \cdots S_m, \cdots \tag{71}$$

In each of these situations we shall have to test the same hypothesis H_0 concerning the probability law $p(x \mid \theta)$ and specifying the value θ_0 of θ. The situations differ among themselves by the number of the X's and by the hypotheses, alternative to H_0, which are considered. For the situation S_m we shall denote them by n_m and H_m respectively. We shall assume that $\lim n_m = +\infty$ when $m \to \infty$. As to the hypothesis H_m, we shall assume that the value θ_m which it ascribes to the parameter θ is

$$\theta_m = \theta_0 + \frac{\vartheta}{\sqrt{n_m}} \tag{72}$$

---

[8] (1) S. Kołodziejczyk: Sur l'erreur de la seconde catégorie dans le problème de "Student." C. R. Academie des Sciences, Paris, t. 197 (1933) p. 814.

(2) J. Neyman with co-operation of K. Iwaszkiewicz and S. Kolodziejczyk: Statistical Problems in Agricultural Experimentation. Suppl. Journ. Roy. Stat. Soc. Vol. II (1935) pp. 107–180.

(3) J. Neyman and B. Tokarska: Errors of Second Kind in Testing "Student's" Hypothesis. J. A. S. A., Vol. 31 (1936) pp. 320–334.

[9] S. S. Wilks and Catherine M. Thompson: The Sampling Distribution of the Criterion λ_H, when the Hypothesis Tested is not true. Biometrika, Vol. XXIX (1937), pp. 124–132.

where ϑ, the standardized error in θ_0, is kept constant. We shall assume that in each situation S_m we test the hypothesis H_0 by means of the test unbiased in the limit and corresponding to the level of significance α. The power function of this test should be denoted by $\beta(\theta \mid \varpi_{n_m})$, but to simplify the notation we will write simply $\beta_m(\theta)$. We shall be concerned with the value of this function $\beta_m(\theta_m)$ at the point $\theta = \theta_m$ and we shall prove the following proposition.

PROPOSITION III. *If the third logarithmic derivative of $p(x_i \mid \theta)$ with respect to θ is bounded*

$$\left| \frac{\partial^3 \log p(x_i \mid \theta)}{\partial \theta^3} \right| < C = \text{constant}, \tag{73}$$

and $|r| < 1$, *then*

$$\lim_{m \to \infty} \beta_m(\theta_m) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}(x - \vartheta\sigma_1)^2} \frac{1}{\sqrt{2\pi}} \int_{M - Nx^2}^{\infty} e^{-\frac{1}{2}y^2}\, dy \right\} dx \tag{74}$$

This proposition is analogous to that[10] concerning the "smooth" test for goodness of fit. It could be used in the following manner.

When testing the hypothesis H_0 and using for the purpose a certain number n of observations, we find ourselves in a situation which might be considered as one of the sequence (71). If n is large, we may hope that the right hand side of (74) will give a reasonable approximation to the actual value of the power function corresponding to the value of θ to be calculated from (72) by substituting in it $n_m = n$.

PROOF. Denote

$$\frac{\partial^3 \log p(x_i \mid \theta)}{\partial \theta^3} = \chi_i(\theta) \tag{75}$$

We may write

$$p(x_i \mid \theta_m) = p(x_i \mid \theta_0) e^{\frac{\vartheta \varphi_i}{\sqrt{n_m}} + \frac{\vartheta^2 \Psi_i}{2 n_m} + \frac{\vartheta^3 \chi(\theta_i')}{6 n_m^{3/2}}} \tag{76}$$

where θ_i' denotes some value intermediate between θ_0 and θ_m. Consequently, taking into account (39), (47) and (48), we have

$$p(E_{n_m} \mid \theta_m) = \prod_{i=1}^{n_m} p(x_i \mid \theta_m) = p(E_{n_m} \mid \theta_0)(1 + \epsilon_m) e^{x\vartheta\sigma_1 - \frac{1}{2}\vartheta^2\sigma_1^2} \tag{77}$$

where

$$\log(1 + \epsilon_m) = \frac{1}{\sqrt{n_m}} \left(\frac{1}{2} \vartheta^2 \sigma_2 (y\sqrt{1 - r^2} + xr) + \vartheta^3 \frac{\sum_{i=1}^{n_m} \chi(\theta_i)}{n_m} \right) \tag{78}$$

---

[10] J. Neyman: "Smooth" Test for Goodness of Fit. Skandinavisk Aktuarietidskrift, (1937), p. 186.

It is seen that, if $m \to \infty$ then ϵ_m tends to zero, uniformly in every bounded region of the plane, S, of x and y. Denote by s any bounded region in S and by $W_m(s)$ a region in W_{n_m} of which s is a transformation by means of the formulae (39), (47) and (48). The probability of E_{n_m} falling within $W_m(s)$ is equal to that of the point with coordinates x and y falling within s. The former of these probabilities is represented by the integral of (77) over $W_m(s)$ and the latter by the integral taken over s of the elementary probability law $p_m(x, y \mid \theta_m)$ of x and y, corresponding to the value θ_m of θ. Owing to the formula (77) we may write

$$(79) \qquad p_m(x, y \mid \theta_m) = p_m(x, y \mid \theta_0)(1 + \eta_m)e^{x\vartheta\sigma_1 - \frac{1}{2}\vartheta^2\sigma_1^2}$$

where, owing to (78), η_m tends uniformly to zero in s as $m \to \infty$. Remembering the connection between u_n, v_n and x, y and also the inequality (56), which is valid for sufficiently large values of n, we conclude that

$$(80) \qquad p_m(x, y \mid \theta_0) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} + Q_m$$

where Q_m has the property that, whatever be $\epsilon > 0$, for sufficiently large values of m

$$(81) \qquad \left| \iint_s Q_m \, dx \, dy \right| < \epsilon$$

where s is any bounded region in S. It follows that

$$(82) \qquad p_m(x, y \mid \theta_m) = \frac{1 + \eta_m}{2\pi} e^{-\frac{1}{2}\{(x-\vartheta\sigma_1)^2+y^2\}} + Q_m e^{x\vartheta\sigma_1 - \frac{1}{2}\vartheta^2\sigma_1^2}(1 + \eta_m)$$

and that therefore, whatever be the bounded region s

$$(83) \qquad \lim_{m\to\infty} \iint_s p_m(x, y \mid \theta_m) \, dx \, dy = \frac{1}{2\pi} \iint_s e^{-\frac{1}{2}\{(x-\vartheta\sigma_1)^2+y^2\}} \, dx \, dy$$

It is known however, that whenever an integral probability law tends to a fixed limit uniformly within any bounded region, then it must do so within the whole space. It follows therefore that the formula (83) is valid for any region s whether bounded or not. But

$$(84) \qquad \beta_m(\theta_m) = \iint_{y > M - Nx^2} p_m(x, y \mid \theta_m) \, dx \, dy$$

and it follows that

$$(85) \qquad \lim_{m\to\infty} \beta_m(\theta_m) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ e^{-\frac{1}{2}(x-\vartheta\sigma_1)^2} \frac{1}{\sqrt{2\pi}} \int_{M-Nx^2}^{+\infty} e^{-\frac{1}{2}y^2} \, dy \right\} dx,$$

which completes the proof of Proposition III.

It is important to be clear about the exact meaning of the Proposition III. Suppose for example that in a particular case $\vartheta = \sigma_1 = 1$ and consider a sequence of situations in which

$$\begin{cases} n_1 = 100, & n_2 = 100^2, & \cdots n_m = 100^m, \cdots \\ \theta_1 = \theta_0 + \cdot 1, & \theta_2 = \theta_0 + \cdot 01, & \cdots \theta_m = \theta_0 + (\cdot 1)^m, \cdots \end{cases} \tag{86}$$

If this were the case, then the Proposition III would be applicable and we could affirm that the sequence of the power functions $\beta_m(\theta)$, each considered at the appropriate point θ_m, has a limit, represented by the double integral in the right hand side of (85) with $\vartheta\sigma_1 = 1$. Accordingly, if we were interested in the value of the power function at $\theta' = \theta_0 + \cdot 02$ with $n = 10000$ and $\theta_1 = 1$, then we could hope to obtain its approximate value calculating the double integral in (85) with

$$\vartheta = (\theta' - \theta_0)\sqrt{n} = 2 \tag{87}$$

These are legitimate conclusions. However, it would be wrong to consider as proved that, if in the same example we increase the size of n to $n' = 40000$, then the value of the power function at $\theta = \theta'$ will be represented by its limit (85) with $\vartheta = 4$ and with about the same accuracy as previously. It is just possible that to attain the same accuracy at $\vartheta = 4$ a value of n greater than n' will be needed. This of course would imply a corresponding change in θ'.

Proposition IV. *If the conditions of Proposition III are satisfied but either* $|r| = 1$ *or* Ψ_i *is independent of* x_i, *then*

$$\lim_{m\to\infty} \beta_m(\theta_m) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\lambda}^{+\lambda} e^{-\frac{1}{2}(x-\vartheta\sigma_1)^2}\,dx \tag{88}$$

The proof of this proposition is quite analogous to that of Proposition III.

6. Examples.

Example 1. Consider the case where it is known for certain that

$$p(x_i|\theta) = \frac{1}{\pi}\frac{1}{1+(x_i-\theta)^2}, \quad \text{for } -\infty < x < \infty \tag{89}$$

but where the actual value of θ is doubtful and it is desired to test the hypothesis H_0 that $\theta = \theta_0 = 0$, the alternative possibilities being both $\theta < 0$ and $0 < \theta$. Before applying the test unbiased in the limit it is natural to try the unbiased test of type A. The critical region w_0 of this test is defined by the inequality

$$\sum_{i=1}^{n} \Psi_i + \left(\sum_{i=1}^{n} \varphi_i\right)^2 \geq a + b\sum_{i=1}^{n} \varphi_i \tag{90}$$

where the constants a and b must be found so as to satisfy the conditions

$$\int \cdots \int_{w_0} p(E_n|\theta_0)\,dx_1 \cdots dx_n = \alpha \tag{91}$$

$$\int \cdots \int_{w_0} \sum_{i=1}^{n} \varphi_i p(E_n|\theta_0)\,dx_1 \cdots dx_n = 0 \tag{92}$$

The technical difficulties involved in this problem are considerable and this may induce us to apply the test unbiased in the limit. Following the above theory we have

(93) $$\varphi_i = \frac{2x_i}{1 + x_i^2}$$

(94) $$\Psi_i = \frac{4x_i^2}{(1 + x_i^2)^2} - \frac{2}{1 + x_i^2}$$

(95) $$\chi_i(\theta) = \frac{16(x_i - \theta)^3}{(1 + (x_i - \theta)^2)^3} - \frac{12(x_i - \theta)}{(1 + (x_i - \theta)^2)^2}$$

It is easily seen that all the limiting conditions of the theory are fulfilled and that, in particular $|\chi_i(\theta)|$ cannot exceed a fixed limit, approximately equal to 3. We have further

(96) $$\mathcal{E}(\varphi_i^2) = -\mathcal{E}(\Psi_i) = \frac{4}{\pi}\int_{-\infty}^{+\infty} \frac{x_i^2}{(1 + x_i^2)^3}\, dx_i = \tfrac{1}{2} = \sigma_1^2$$

Similarly

(97) $$\mathcal{E}\{\Psi_i + \sigma_1^2)^2\} = \tfrac{5}{8} = \sigma_2^2$$

(98) $$\mathcal{E}(\varphi_i \Psi_i) = 0 = r$$

It follows that the regions $\bar{w}_n$, the sequence of which determines the test which is unbiased in the limit, are defined by the inequality

(99) $$4\sum_{i=1}^{n} \frac{x_i^2}{(1 + x_i^2)^2} - 2\sum_{i=1}^{n} \frac{1}{1 + x_i^2} + 4\left(\sum_{i=1}^{n} \frac{x_i}{1 + x_i^2}\right)^2 \geq M\sqrt{\frac{5n}{8}} - \frac{n}{2}$$

where M should be calculated so as to satisfy (27) with

(100) $$N = \sqrt{\frac{2n}{5}}$$

In order to test the hypothesis H_0 we have therefore to observe the values $x_1, x_2, \cdots x_n$ and to substitute them into the left hand side of (99). If the inequality is satisfied then the hypothesis should be rejected.

Approximate values of the power function could be obtained from the right hand side of (85) with

(101) $$\vartheta\sigma_1 = \theta\sqrt{\frac{n}{2}}.$$

EXAMPLE II. Let us assume as given that

(102) $$\begin{cases} p(x_i|\theta) = \theta e^{-\theta x_i} & \text{for } 0 < x_i \\ \quad = 0 & \text{elsewhere} \end{cases}$$

with $\theta > 0$, the hypothesis to test being that $\theta = \theta_0 = 1$, with the alternatives both $\theta < 1$ and $\theta > 1$.

In this particular example the unbiased test of type A is easily found[11] and moreover[12] it has also the property of being of type A_1. But this circumstance does not diminish the illustrative character of the example. We have

$$\varphi_i = 1 - x_i \tag{103}$$

$$\Psi_i = -1 = \text{constant} \tag{104}$$

It follows that the regions forming the test which is unbiased in the limit are determined by the inequality (59). We have further

$$\sigma_1^2 = \mathcal{E}(\varphi_1^2) = \int_0^\infty (1 - x^2)e^{-x}\,dx = 1 \tag{105}$$

and the inequality (59) reduces to

$$\left|\sum_{i=1}^{n} (1 - x_i)\right| \geq \lambda\sqrt{n} \tag{106}$$

with λ taken from the tables of the normal integral according to (60) and to the chosen value α. Approximate values of the power function can be calculated from, say

$$\beta_\infty(\vartheta) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\lambda}^{+\lambda} e^{-\frac{1}{2}(x-\vartheta)^2}\,dx \tag{107}$$

with

$$\vartheta = (\theta - 1)\sqrt{n} \tag{108}$$

The simplicity of the example considered permits to calculate the exact power function of the test and it may be interesting to obtain its limit $\beta_\infty(\vartheta)$ in another and a more direct way. Write

$$\sum_{i=1}^{n} x_i = y \tag{109}$$

It is known that, if the probability law of each of the X's is given by (102) then the probability law of y is

$$p(y \mid \theta) = \begin{cases} \dfrac{\theta^n y^{n-1}}{(n-1)!} e^{-\theta y} & \text{for } 0 < y \\ = 0 & \text{otherwise} \end{cases} \tag{110}$$

---

[11] J. Neyman and E. S. Pearson, loc. cit. p. 18 et seq.

[12] J. Neyman: Estimation statistique traitée comme un problème de probabilité classique. Series Actualités scientifiques et industrielles. Paris, (1938). (In the press.)

It follows that the exact form of the power function corresponding to the test (106) is

$$(111) \qquad \beta(\theta \mid \overline{w}_n) = 1 - \frac{\theta^n}{(n-1)!} \int_{n-\lambda\sqrt{n}}^{n+\lambda\sqrt{n}} y^{n-1} e^{-\theta y}\, dy$$

For values of n about 100 or more and for the values of θ close to unity the distribution of say

$$(112) \qquad z = \frac{\theta \sum x_i - n}{\sqrt{n}} = \frac{\theta y - n}{\sqrt{n}}$$

is practically normal with mean equal to zero and S.D. equal to unity. It follows that the integral in the right hand side of (111) is practically equal to the normal integral taken within the limits which are obtained by substituting in (112) the limits of y in (111). After some easy transformation we have, with a considerable accuracy

$$(113) \qquad \beta(\theta \mid \overline{w}_n) = 1 - \frac{1}{\sqrt{2\pi}} \int_{(\theta-1)\sqrt{n}-\lambda\theta}^{(\theta-1)\sqrt{n}+\lambda\theta} e^{-\frac{1}{2}z^2}\, dz$$

or, after some further transformations and taking into account (108)

$$(114) \qquad \beta(\theta \mid \overline{w}_n) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\lambda(1+\vartheta/\sqrt{n})}^{+\lambda(1+\vartheta/\sqrt{n})} e^{-\frac{1}{2}(u-\vartheta)^2}\, du$$

and it is seen that, when ϑ is fixed and n indefinitely increases, then $\beta(\theta \mid \overline{w}_n)$ does tend to $\beta_\infty(\vartheta)$.

UNIVERSITY COLLEGE, LONDON.

THE TRANSFORMATION OF STATISTICS TO SIMPLIFY THEIR DISTRIBUTION*

BY HAROLD HOTELLING AND LESTER R. FRANKEL

1. Introduction. The custom of regarding a result as significant if it exceeds two or three times its standard error has now given way among informed statisticians to a consideration of the exact probabilities associated with the distribution of the statistic in question. For example, in such problems as that of examining the significance of the difference between the means of two samples, particularly small samples, it is no longer adequate to regard the difference of means, divided by the sample estimate of its standard error, as normally distributed. The significance of this ratio, "Student's ratio," is judged instead by the value of

$$P = 2 \int_t^\infty \phi_n(z)\, dz \tag{1}$$

where n is the number of degrees of freedom entering into the estimate of variance, and

$$\phi_n(z) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{z^2}{n}\right)^{\frac{1}{2}(n+1)}}. \tag{2}$$

If the probability law underlying the observations themselves is normal, and they are independent, P is the exact probability of the value of t obtained being equalled or exceeded on the hypothesis that there is no real difference between the means.

Methods of approximating P have been studied by R. A. Fisher[1] and by W. A. Hendricks,[2] and tables have been presented by Student[3] and Fisher.[4] Nevertheless, the practical statistician will very frequently wish to make judgments of significance without stopping to consult a table, or laboriously to compute P, and will tend to revert to the former inaccurate but convenient practice of treating t as normally distributed with unit variance. The essential

* Presented at the joint meeting at Indianapolis of the American Mathematical Society and the Institute of Mathematical Statistics, December 30th, 1937.

[1] *Expansion of Student's Integral in Powers of n^{-1}.* Metron, vol. 5 (1925).

[2] Annals of Mathematical Statistics, vol. 7 (1936), pp. 210-221.

[3] *New Tables for Testing the Significance of Observations.* Metron, vol. 5 (1925).

[4] *Statistical Methods for Research Workers,* Oliver and Boyd, 1925-1936. Tables IV and VI.

reason for this is that the normal distribution to which that of t approximates for large values of n has only one parameter in the expression for the probability. Hence it is easy to remember a few important values, such as those corresponding to $P = .01$ and $.05$; and when values of P representing other levels of significance are in question, the single-entry tables of the normal probability integral are more easily available and easier to use than the double-entry table of Student's integral. Indeed, t is a more useful statistic than Student's original ratio of mean to sample standard deviation, to which it is in the simplest case proportional, partly because of the close approximation of t for large samples to a normally distributed variate of unit variance.

For more complicated statistics the practical need for something simpler than the exact distribution is even more urgent, on account of the larger number of parameters involved in the distributions. For example, the large class of problems giving rise to probabilities expressible as incomplete beta functions require for exactitude the use of Pearson's extensive triple-entry table,[5] and even this is inadequate for some ranges of the parameters. The shorter tables of R. A. Fisher[6] and of Snedecor[7] are helpful, but are also necessarily of triple entry.

It is a common practice, for example, among economists and psychologists, to select either by graphic methods or by preliminary calculation that one, out of many tests that might be applied to available data, for which P is the least. Such selection evidently introduces a bias, which is the more subtle because the tests giving high and therefore insignificant probabilities are likely to be forgotten. Often the only way to guard against such fallacies is to insist on a value of P lower than is easily determined from tables. Thus, if k independent tests of significance have been made, and only the smallest value P is reported, its significance should be judged not by this value P itself, but by the probability

$$P' = 1-(1 - P)^{k} = kP - \cdots$$

of the least value being so small. If we equate P' to some such standard value as .01, then P must, for this standard level of confidence, take only a fraction, approximately $1/k$, of this value. Such a small probability will often fall outside the range of existing tables.

Instead of relying on tables or direct computation from the exact distribution of a statistic, it will sometimes be desirable to use a modification of the statistic, selected so as to have the normal or some other standard distribution. We shall consider a type of transformation of a statistic such that the distribution becomes the limiting form of the original distribution as the sample size increases. Thus our transformation will reduce to the application to the statistic of a correction which will be small when the sample is large. We shall show how to make simple approximate corrections of this character for two cases.

---

[5] *Tables of the Incomplete Beta Function*, Biometrika Office, 1934.

[6] Loc. cit. Tables IV and VI.

[7] *Calculation and Interpretation of Analysis of Variance and Covariance.* Ames, Iowa. Collegiate Press. 1934.

The first of these is the Student ratio t, the lower limit of the integral in (1). Putting

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{3}$$

and

$$P = 2 \int_x^\infty \phi(z)\, dz \tag{4}$$

which in view of (1) and the fact that the integral of each distribution from $-\infty$ to ∞ is unity is equivalent to

$$\int_0^x \phi(z)\, dz = \int_0^t \phi_n(z)\, dz \tag{5}$$

we shall show that x has an asymptotic expansion:

$$x \sim t\left\{1 - \frac{t^2+1}{4n} + \frac{13t^4 + 8t^2 + 3}{96n^2} - \frac{35t^6 + 19t^4 + t^2 - 15}{384n^3} + \frac{6271t^8 + 3224t^6 - 102t^4 - 1680t^2 - 945}{92{,}160n^4} - \cdots\right\} \tag{6}$$

It will frequently be a sufficient approximation to treat

$$t\left(1 - \frac{t^2+1}{4n}\right)$$

as normally distributed. These appear to be approximations of practical value when $n \geq t^2$.

The second statistic whose transformation to a function having its limiting distribution we shall consider is the generalized Student ratio T, appropriate to all the uses to which t can be put, but with a multiplicity of variates instead of one to serve as the basis of the test of significance.[8] This is defined with reference to variates $x_1, \cdots x_p$, together with a linear function of sample values (proportional for example to the difference between the means in two samples), such that if ξ_i is the value of this function of the sample values of x_i ($i = 1, \cdots, p$) then the variance of ξ_i in the population sampled is the same as that of x_i, and on the hypothesis to be tested, the population mean of each ξ_i is zero. In terms of unbiased quadratic estimates s_{ij} of the covariances σ_{ij} among $x_1, \cdots, x_p$, each based on n degrees of freedom, we may define l_{ij} as the cofactor of s_{ij} divided by the determinant of the statistics s_{ij}. Then T is defined by

$$T^2 = \Sigma\Sigma l_{ij}\xi_i\xi_j \tag{7}$$

[8] Harold Hotelling, *The Generalization of Student's Ratio.* Annals of Mathematical Statistics, vol. 2 (1931), pp. 360–378.

the summations running independently with respect to i and j from l to p. For independent samples from a multivariate normal population, the distribution of T has been shown[9] to be

$$\frac{2\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{n-p+1}{2}\right)n^{\frac{1}{2}p}} \frac{T^{p-1}\,dT}{\left(1+\frac{T^2}{n}\right)^{\frac{1}{2}(n+1)}}. \tag{8}$$

As n increases, the distribution of T approaches the χ distribution with p degrees of freedom:

$$\frac{\chi^{p-1}e^{-\frac{1}{2}\chi^2}\,d\chi}{2^{\frac{1}{2}(p-2)}\,\Gamma\left(\frac{p}{2}\right)} \tag{9}$$

By equating the probabilities derived from these two distributions, we shall define χ as a function of T, and obtain asymptotic expansions for the functions χ and χ^2 thus defined.

Since the probability associated with T is expressible in terms of the incomplete beta function, or the analysis of variance distribution integral, it follows that any of the many common statistics, of which simple functions have this distribution, can be expressed simply in terms of T. Tests of significance in a wide variety of cases may therefore be made with the help of the asymptotic expansion corresponding to T^2, together with a table of χ^2.

A further advantage of the transformation of a statistic into a normally distributed variate of unit variance and zero mean is that further statistical tests are possible with such variates. Since a great part of statistical theory is based on the assumption of such normal distributions, an extensive field of applications becomes available in this way. For example, if several independent tests give values of t based on various numbers of degrees of freedom, and it is desired to combine these tests so as to get a single probability, the corresponding values of the normally distributed variate x defined above may be squared and added. The sum will then have the χ^2 distribution, with a number of degrees of freedom equal to the number of values of t used. In a similar manner, the values of χ^2 corresponding to a number of independently determined values of T^2 may be added, and the sum will have the χ^2 distribution with a number of degrees of freedom equal to the sum of the various values of p involved.

The advantages of this type of what may be called "normalization" of a statistic have been brought out by R. A. Fisher for the particular case of the correlation coefficient. His use[10] of $z = \frac{1}{2}\log\frac{1+r}{1-r}$ facilitates such operations as the averaging of values obtained from independent samples, or taking the

---

[9] Harold Hotelling, loc. cit.

[10] *Statistical Methods for Research Workers*, Sec. 35.

difference between two values, with the testing of significance of the result in each case. This is because z, unlike r, has a nearly normal distribution, with variance nearly independent of the population value. We note in passing that this function is the same as $\tanh^{-1} r$, and may therefore be determined accurately and readily from the Smithsonian Institution Tables of Hyperbolic and Exponential Functions.

2. **Normalization of t.** The "duplication formula" in the theory of the Gamma function[11] shows that

$$\Gamma\left(\frac{n+1}{2}\right) = \frac{\sqrt{\pi}\,\Gamma(n)}{2^{n-1}\Gamma\left(\frac{n}{2}\right)}$$

Substituting this in (2) and taking logarithms we have:

$$\begin{aligned}\log \phi_n(z) = -\tfrac{1}{2}\log n - (n-1)\log 2 + \log \Gamma(n)\\ - 2\log\Gamma\left(\frac{n}{2}\right) - \frac{n+1}{2}\log\left(1+\frac{z^2}{n}\right)\end{aligned} \tag{10}$$

The last logarithm may be expanded in a series of powers of z^2/n which not only converges uniformly on the interval $0 \leq z \leq t$ when $n > t^2$, but has the property of being a *uniformly asymptotic* representation of the function on this interval. This means that the sum of the first j terms of the series ($j = 0, 1, 2, \cdots$) differs from the function represented, by a quantity whose product by n^{j+1} has, for sufficiently large values of n, an upper bound independent of z, so long as z remains in this interval. Uniformly asymptotic series have a number of important properties, among which is[12] term by term integrability with respect to z. In this sense we have the uniform asymptotic representation:

$$-\frac{n+1}{2}\log\left(1+\frac{z^2}{n}\right) \sim -\frac{z^2}{2} - \frac{2z^2 - z^4}{4n} + \frac{3z^4 - 2z^6}{12n^2} - \frac{4z^6 - 3z^8}{24n^3} + \cdots \tag{11}$$

We shall obviously have another uniform asymptotic representation if we add to this, term by term, asymptotic series with terms independent of z, such as those for the gamma function logarithms in (10). Since[13]

$$\log\Gamma(n) \sim \tfrac{1}{2}\log 2\pi + (n - \tfrac{1}{2})\log n - n + \sum_{r=1}^{\infty}\frac{(-1)^{r-1}B_r}{2r(2r-1)n^{2r-1}}, \tag{12}$$

where

$$B_1 = \tfrac{1}{6}, \qquad B_2 = \tfrac{1}{30}, \qquad B_3 = \tfrac{1}{42}, \qquad B_4 = \tfrac{1}{30}, \qquad B_5 = \tfrac{5}{66}, \cdots$$

[11] Whittaker and Watson, *Modern Analysis*, 4th ed., p. 240.

[12] H. Schmidt, *Beiträge zu eine Theorie der allgemeinen asymptotischen Darstellungen.* Math. Annalen, vol. 113 (1937), pp. 629–656. The property mentioned above is proved in Schmidt's Theorem 6.

[13] Whittaker and Watson, loc. cit., pp. 252, 125.

are the Bernoulli numbers, we obtain upon substituting in (10) this and the similar formula for $\log \Gamma\left(\frac{n}{2}\right)$, together with (11), and some simplification,

$$\begin{aligned} \log \phi_n(z) \sim -\tfrac{1}{2}\log 2\pi - \frac{z^2}{2} + \frac{-1 - 2z^2 + z^4}{4n} \\ (13) \qquad + \frac{3z^4 - 2z^6}{12n^2} + \frac{1 - 4z^6 + 3z^8}{24n^3} + \frac{5z^8 - 4z^{10}}{40n^4} + \cdots \end{aligned}$$

Upon differentiating (5) we obtain:

$$(14) \qquad \phi(x)\frac{dx}{dt} = \phi_n(t)$$

Since ϕ is simply the normal distribution function (3), this may be written:

$$(15) \qquad -\frac{1}{2}\log 2\pi - \frac{x^2}{2} + \log\frac{dx}{dt} = \log \phi_n(t)$$

We shall always in this paper use the symbol "lim" to mean the limit as n approaches infinity. The functions of n and z, or of n and t, which we shall denote by R, R', R'', with or without subscripts, are to be such that the absolute value of each has an upper bound independent of n, z and t so long as $n \geq 1$, and z and t are confined to some fixed finite interval.

From (13) we have that $\lim \log \phi_n(z) = \log \phi(z)$, whence, by the continuity of the exponential function,

$$\lim \phi_n(z) = \phi(z)$$

This holds *uniformly* for $0 \leq z \leq t$. Subtracting $\int_0^t \phi(z)\,dz$ from both sides of (5) we therefore find that

$$(16) \qquad \int_t^x \phi(z)\,dz = \int_0^t \{\phi_n(z) - \phi(z)\}\,dz$$

can by choosing n large enough be made as small as we please. Since $\phi(z) > 0$, it follows that the function x of t and n is such that

$$(17) \qquad \lim x = t.$$

A parallel argument, proving slightly more than (17), is the following. From (13),

$$\log \phi_n(z) = \log \phi(z) + \frac{R'}{n}$$

where R' is a bounded function of the kind described above. Therefore

$$\phi_n(z) = \phi(z)\left(1 + \frac{R''}{n}\right).$$

Substituting this in (16) we have that

$$\int_t^x \phi(z)\,dz = \frac{1}{n}\int_0^t \phi(z)R''\,dz$$

From the mean value theorem of integral calculus it then follows that

$$x = t + \frac{R_1}{n} \tag{18}$$

An asymptotic series may be substituted in a power series, and the result is a valid asymptotic representation of the corresponding function. (Schmidt, loc. cit., Theorem 4.) This justifies taking the exponential of each side of (13) and arranging in a series of powers of n^{-1} to give

$$\phi_n(z) \sim \phi(z)\left\{1 + \frac{\alpha_1(z)}{n} + \frac{\alpha_2(z)}{n^2} + \cdots\right\} \tag{19}$$

This asymptotic development will, like the original one, hold uniformly in every finite interval, and may therefore be integrated term by term. Thus

$$\int_0^t \phi_n(z)\,dz = \int_0^t \phi(z)\left\{1 + \frac{\alpha_1(z)}{n} + \frac{\alpha_2(z)}{n^2} + \cdots + \frac{\alpha_j(z)}{n^j}\right\}dz + \frac{R_{j+1}}{n^{j+1}} \tag{20}$$

where $|R_{j+1}|$ has an upper bound independent of n and t when $n \geq 1$, and t is confined to a finite interval, $0 \leq t \leq T$. Substituting this in (16) we obtain:

$$\int_t^x \phi(z)\,dz = \int_0^t \phi(z)\left\{\frac{\alpha_1(z)}{n} + \cdots + \frac{\alpha_j(z)}{n^j}\right\}dz + \frac{R_{j+1}}{n^{j+1}} \tag{21}$$

In terms of a sequence of functions $f_1, f_2, \cdots$ of t to be defined below, let

$$x_j = t + \frac{f_1}{n} + \frac{f_2}{n^2} + \cdots + \frac{f_j}{n^j}. \tag{22}$$

Now $\int_t^{x_j} \phi(z)\,dz$ can be expanded in a series of powers of n^{-1} which converges for sufficiently large values of n; for the Taylor series

$$\phi(z) = \phi(t) + (z - t)\,\phi'(t) + \cdots \tag{23}$$

can be integrated to give a series of powers of $x_j - t$, which by (22) is a polynomial in n^{-1}. As a matter of fact we have from (22) that $x_j - t$ can be made arbitrarily small by taking n large enough; consequently the series (23) and that obtained by integration in this way will converge uniformly and absolutely. We thus have:

$$\begin{aligned}\int_t^{x_j} \phi(z)\,dz &= \frac{1}{n}f_1\phi + \frac{1}{n^2}\left(f_2\phi + \frac{1}{2}f_1^2\phi'\right)\\ &\quad + \frac{1}{n^3}\left(f_3\phi + f_1f_2\phi' + \frac{1}{6}f_1^3\phi''\right) + \cdots\end{aligned} \tag{24}$$

Now let us define $f_1, f_2 \cdots$, by equating the coefficient of each power of n in (24) to that of the same power of n in the right member of (21). This process gives a sequence of equations

$$
\begin{aligned}
f_1\phi &= \int_0^t \phi(z)\alpha_1(z)\,dz \\
f_2\phi + \tfrac{1}{2}f_1^2\phi' &= \int_0^t \phi(z)\alpha_2(z)\,dz \\
f_3\phi + f_1f_2\phi' + \tfrac{1}{6}f_1^3\phi'' &= \int_0^t \phi(z)\alpha_3(z)\,dz \\
f_4\phi + (f_1f_3 + \tfrac{1}{2}f_2^2)\phi' + \tfrac{1}{2}f_1^2f_2\phi'' + \tfrac{1}{24}f_1^4\phi''' &= \int_0^t \phi(z)\alpha_4(z)\,dz
\end{aligned}
\tag{25}
$$

Since $\phi \neq 0$ the first of these equations defines f_1 for every value of t; when f_1 has been determined, the second equation defines f_2; then the third defines f_3, and so forth. It is to be observed that the functions $f_1, f_2, \cdots$ thus determined are not changed when the value of j appearing in (22) is increased; we have a unique sequence.

If for the right-hand member of (15) we substitute that of (13), replacing z by t, and on the left of (15) put

$$x = t + \frac{f_1}{n} + \frac{f_2}{n^2} + \cdots$$

$$\frac{dx}{dt} = 1 + \frac{f_1'}{n} + \frac{f_2'}{n^2} + \cdots$$

and then expand in a formal manner in powers of n^{-1}, we shall upon equating coefficients of like powers of n obtain a sequence of differential equations

$$
\begin{aligned}
f_1' - tf_1 &= \tfrac{1}{4}(-1 - 2t^2 + t^4) \\
f_2' - tf_2 &= \tfrac{1}{2}f_1'^2 + \tfrac{1}{2}f_1^2 + \tfrac{1}{4}t^4 - \tfrac{1}{6}t^6 \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
\tag{26}
$$

These, with the initial condition $f_1 = f_2 = \cdots = 0$ for $t = 0$ determine the same sequence of functions as before. The equations (26) are in fact obtainable simply by differentiating (25) and cancelling out the factor $\phi(t)$. That this must be true follow from the equivalence of the various formal processes of manipulating series of powers of n^{-1}, whether convergent or divergent, to give equivalent results. The differential equations are easily solved; the solutions, at least for f_1, f_2, f_3, and f_4, are all polynomials. Why they should come out as polynomials is not immediately obvious; but their calculation is made easier if each f_j is replaced in the differential equations by a polynomial of degree $2j + 1$ with undetermined coefficients, involving only odd powers of t. The f's of lower order are replaced by values previously determined, and the coefficients are

found by equating like powers of t. This process supplies at each stage more equations than unknown coefficients; their consistency verifies the assumption that f_j is a polynomial of the kind specified, at least for $j \leq 4$. These polynomials are the coefficients of the powers of n^{-1} in (6).

The series on the right of (24) not only converges but is an asymptotic series uniformly valid when t varies in any finite interval. Hence upon subtracting (24) from (21) and taking account of (25) we find that

$$\int_{x_j}^{x} \phi(z)\, dz = \frac{R'_{j+1}}{n^{j+1}}$$

where $| R'_{j+1} |$ is uniformly bounded. Upon applying the mean value theorem to the integral on the left we find that x differs from x_j, and thus from the first j terms (22) of the series (6), by a quantity whose product by n^{j+1} remains bounded when n approaches infinity. This proves the validity of the asymptotic expansion.

3. **Accuracy of the Approximation.** To follow through the above processes in such a way as to obtain useful limits for the error involved in using the first few terms of the series (6) in place of x would be excessively difficult. However, the magnitude of the error in taking the first two or three terms as an approximation to x may be judged from the tables below to be adequately small for practical purposes, provided $n \geq t^2$. The essential singularity of the normal distribution at infinity, in contrast with the algebraic nature of the Student distribution, means a poorer approximation of one to the other as t increases while n remains fixed, though a better approximation as n increases. This is illustrated in the following tables, where it will be observed that the approximations are better for large than for small values of n, and of

$$P = 2 \int_{x}^{\infty} \phi(z)\, dz = 2 \int_{t}^{\infty} \phi_n(z)\, dz$$

It will be seen that for $n = 10$ and $P < .001$, the utility of the asymptotic series, or at least of its first five terms, is vitiated by the rapid oscillation of consecutive terms, due to the high values of t^2 in relation to n.

	$P = .10$		$P = .05$		$P = .01$	
	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 10$	$n = 30$
t	1.812	1.697	2.228	2.042	3.169	2.750
x_1	1.618	1.642	1.896	1.954	2.294	2.554
x_2	1.650	1.645	1.980	1.960	2.754	2.579
x_3	1.643	1.645	1.953	1.960	2.446	2.575
x	1.645		1.960		2.576	

	$P = .001$		$P = .0001$		
	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 100$
t	4.587	3.646	6.22	4.482	4.052
x_1	2.059	3.212	.05	3.69	3.88
x_2	4.981	3.313	12.86	3.98	3.89
x_3	0.896	3.283	−20.44	3.85	3.89
x_4	7.163	3.293	75.66	3.91	3.89
x	3.291		3.891		

4. **Transformation of the Generalized Student Ratio.** The arguments and methods of calculation set forth in Section 2 may be applied with little or no change to the transformation of various other statistics in such a way that the limiting distribution for large samples is reached at once for the transformed statistic. In particular, to deal with the generalized Student ratio T, we may equate (8) to (9), represent χ as an asymptotic expansion with undetermined coefficients which are functions of T, and then by substituting and equating like powers of n^{-1} obtain as before a sequence of differential equations for determining the coefficients. This process gives

$$(27) \qquad \chi \sim T - \frac{pT + T^3}{4n} + \frac{(8 - 5p^2)T + (4 + 4p)T^3 + 13T^5}{96n^2} \cdots$$

This reduces to the expansion of x in terms of t previously found if we put $p = 1$.

It is somewhat more convenient in practice to use χ^2 and T^2, to avoid extracting the square root of the latter expression, and to utilize the existing tables of χ^2. Ordinarily therefore we should not use (27), but the series

$$\chi^2 \sim T^2\left\{1 - \frac{p + T^2}{2n} + \frac{(4 - p^2) + (2 + 5p)T^2 + 8T^4}{24n^2} \cdots\right\},$$

which may be obtained in the same way, or by squaring (27) in a formal manner. That these are genuine asymptotic approximations follows by essentially the same argument as before.

Columbia University and Washington, D. C.

ON COMBINED EXPANSIONS OF PRODUCTS OF SYMMETRIC POWER SUMS AND OF SUMS OF SYMMETRIC POWER PRODUCTS WITH APPLICATIONS TO SAMPLING (Continued)

By Paul S. Dwyer

PART II. THE FUNDAMENTALS OF SAMPLING

Introduction

We consider a population of N variates in which every individual possesses a common attribute. Let the variate x_i be the measure of such an attribute for individual i. From the N variates it is possible to form $\binom{N}{n}$ different samples where each sample consists of n variates, $n \leq N$.

Each sample has its mean, variance, etc. so that there are $\binom{N}{n}$ means, $\binom{N}{n}$ variances, etc. The fundamental sampling problem, as interpreted here, is to find the relation between the moments of the $\binom{N}{n}$ means, and the moments of the $\binom{N}{n}$ variances in terms of the moments of the moments of the universe. Numerous attempts have been made to solve this problem, but each has been restricted in some way. It is the aim of Part II to indicate an approach which is broad enough to include many of the fundamental variations.

The first chapter is devoted to a listing of criteria which should be satisfied by a theoretical development which is to be considered sufficiently general. These criteria might be applied to other statistics but the theory developed here is limited to those statistics which are moments (or functions of moments) of moments. The first chapter continues with an account of the more significant papers which have contributed to a general solution of the problem. No attempt is made to indicate a complete history, but rather there is presented a brief summary of a number of the most significant contributions.

The second chapter is devoted to definitions and notation. An attempt has been made to use conventional notation whenever it is suitable.

The third chapter deals with some of the fundamental principles which are used in the general approach. It presents a crucial part of the argument as it shows how various types of sampling problems can be reduced to Carver functions.

The last three chapters deal with specific applications to some of the simpler problems. Chapter IV discusses the case of moments of the mean of the sample.

Chapter V considers the mean of the variance and the variance of the variance, while Chapter VI gives a large number of formulas, implicitly, in tabular form.

Chapter I. A Brief History of Previous Contributions

In order to assist the reader in getting a perspective with reference to previous mathematical work on the relations between the moments of the moments of the sample and the moments of the complete set of measures (universe), a list of criteria[1] is suggested below which might be applied to each contribution. These criteria group themselves naturally into two classes. The first eight questions can be answered categorically, while the remainder are less definite in nature and are not so subject to categorical answers.

1. **The Criteria.** 1. Does the method apply to one type of frequency distribution only or is it broad enough in scope to include any distribution law?

2. Is there any restriction as to the size of the sample?

3. Is there any restriction as to the size of the universe?

4. Is there any restriction as to the nature of the correlation between observations? More specifically, is the method applicable only to some particular law of formation of the sample such as "drawing with replacements," "drawing without replacements," etc., or is it broad enough in scope to allow application to other orderly replacement laws?

5. Is the application limited to one characteristic (variable) or can a large number of characteristics be treated simultaneously?

6. Is it necessary that the universe maintain the same frequency distribution during the formation of the sample or may it assume a different frequency distribution before each drawing?

7. Does the method produce exact, rather than approximate, formulas?

8. Does the method permit approximations to a required degree of accuracy?

9. Does the method enable the author to write general laws in a compact form? More specifically, can he express, in a form which is not too symbolic, any moment of a given sample moment? If not, what order of moments can be expressed?

10. Is the notation such that the general case can be turned into the more important special cases with relative ease?

11. Does the development lead logically to the introduction of new moment functions (such as the semi-invariant of Thiele [B'; 209] or the k functions of R. A. Fisher [23; 203]) which are useful in condensing the results?

12. Is a combinatorial analysis provided so that any given formula, or any part of it, can be checked for accuracy without too much effort?

2. **Review of previous results.** The articles below have been examined with the criteria in mind. No attempt is made to write specific answers to all the

---

[1] Many of these criteria have been suggested, in less explicit form by Tchouproff (15; 461–471). The "Introduction" of his Metron paper is recommended for use as a supplement to the present chapter.

criteria in each case, but rather to indicate the important features of each contribution.

The papers discussed by no means cover all the work on moments of moments, although a rather complete bibliographical background is available to the reader who desires to examine the bibliographies attached to the articles mentioned. Undoubtedly the importance of the articles written in English has been overemphasized. Since the important contributions of non-English writers (such as Thiele and Tchouproff) have eventually appeared in English, it does no serious harm to refer to the English versions even though the results may have been partially antedated by the author in some other language.

A large number of the earlier results on moments of moments were limited to a special case of the problem, usually the case in which the universe is infinite and normal. The present summary deals with those authors who, during the past four decades, have made real contributions to the problem of generalization. A detailed account of the history of moments of moments would include many valuable contributions which are not included here.

It seems expedient to start with Pearson's article "On the Probable Error of Frequency Constants" [2] which appeared at the opening of the century. Although by no means the first article in the field, it presented a rather complete set of formulas for the case of moments of moments. One advantage of these formulas is that they are relatively brief and yet this brevity results from the fact that they are approximate. The original paper dealt with the univariate case, but it was followed by a later one [6] which discussed the case of more than one variable.

These formulas have played an important rôle in that they have assisted in making it clear that the moments of moments of samples must be estimated if one is to be permitted to draw conclusions from his sampling moments and that it is possible to work out formulas which serve as the basis of those estimates.

Of great importance also was the contribution of T. N. Thiele to the sampling problem. Adapting certain ideas of Laplace, he used semi-invariants in which to express his results which he published in English in 1903 in "The Theory of Observations" [B'; 209]. He took the case of the infinite parent and any law of distribution and then worked out moments through the fourth of the variance.

An earlier contribution of the introductory period was that of Karl Pearson in 1899 [1]. This paper is significant in that it provides formulas for the four moments of the mean when sampling is from a *finite* universe. The universe is not general, but obeys a simple frequency law.

Another article of this period was that of Robert Henderson (1904) in which the first four moments of the mean were given for an infinite universe with any frequency law. This article, which was first published in *Transactions of the Actuarial Society of America* [3], was considered so important that it was republished in 1907 in the *British Journal of the Institute of Actuaries.* Henderson gave, in addition to the first four moments of the mean, first moments of m_2, m_3, m_4 although the last of these formulas is erroneous.

Another important contribution of this period was that of "Student" in 1908 [5]. He was interested in the properties of the normal distribution, but did not assume normality in his general derivation. He took an infinite population and wrote the formula for the variance of the variance. In this result he inserted the condition for normality. His further argument in the normal case implied the development of corresponding formulas for the higher moments of the variance, but he did not publish them as they were incidental to his main attack. The semi-invariant equivalent of these results had been previously given by Thiele [B'; 209–210].

The real contribution of "Student" to the general problem of moments of moments was his method, for it is his method which has been utilized by later writers. "Student's" method has the advantage that the development involves algebraic processes only. Contributions of Neyman, Church, Pepper, Carver, and the present writer are based upon it.

An important development during the next decade, 1908–1918, was the establishment of the first four moments of the mean when the samples were drawn from a finite parent without replacement. It appears that a number of men worked this problem independently. For example, one might examine the results of Pearson [4], Isserlis [7, 8], Mortara [C], Tchouproff [11], and Edgeworth [9]. Probably the best English presentations of that era were those of Isserlis [8] and Edgeworth [9] which appear in the same volume of the *Journal of the Royal Statistical Society*.

A most prolific writer on sampling during the next decade was the Russian, Tchouproff, who had been publishing in Russian and Scandinavian journals [10], [11]. His most valuable contributions were published in 1918–1923 in *Biometrika* (in English) and in *Metron* (in English).

The first series of articles was published in three different numbers of *Biometrika* in the years 1918–19 [12]. Tchouproff assumed an infinite universe and used the method of mathematical expectation. At first glance the most characteristic aspect of his work appears to be the complicated notation which he used. This notation was adopted because he undertook a much more general problem than had previously been attempted and hence needed to make new distinctions. Although he limited himself to the infinite case and one variable, he worked out the theory with the freedom that the frequency distribution of the universe might change between drawings. In the special case in which the populations are the same, he worked out the moments of the variance as far as the fourth. The chief criticism of his work concerns the complicated notation which seems to have been difficult to follow critically. A mistake in one of his formulas was not discovered for some years and then not by examination of his reasoning, but through the application of his results to an actual problem [17].

It is perhaps appropriate to insert here that in 1934 Feldman [30] rewrote the material of the second *Biometrika* article by simplifying the notation and extending the argument to the case of two (and more) variables.

Tchouproff continued to generalize his work and in the 1923 volume of *Metron*

[15] there appeared a series of articles in which there were no restrictions as to the size of the sample, no restrictions as to the type of sampling distribution (in fact the sampling distribution might vary between successive drawings), and no restrictions as to the law of replacement, or more generally as he expressed it, "no restriction as to the nature of the correlation between observations." Criterion number 5 is the only one of the first eight criteria which is not satisfied in as much as the approach is limited to that of a single variable. Also the notation was extremely complicated and, although Tchouproff gave general formulas for moments of moments, these formulas are so symbolic in form that he did not find it expedient to write out specific formulas beyond the variance of the variance for such an important special case as sampling from a finite parent without replacements.

During the same period J. Splawa-Neyman [14] had been examining the problem of sampling from a finite parent without replacements. He published his results in a Polish journal in 1923 [14] and his corrected results two years later in *Biometrika* [18]. He gave the well known formulas for the first four moments of the mean and a formula for the variance of the variance. He also gave some simple correlation formulas such as the correlation between the mean and the variance.

At this time the basic problem of moments of moments, at least as it was interpreted by Pearson and his followers, was the establishment of the first four moments of the given moment of the sample so that a Pearson curve could be fitted. A. E. R. Church, a worker in Pearson's laboratory, was assigned the task of seeing how the moments of the variance work out in actual practice. In doing this he became convinced that the formula for the fourth power of the variance, which had appeared in Tchouproff's *Biometrika* article, was incorrect. He tried to follow the argument of Tchouproff, but apparently was baffled by the complex notation and finally, at the suggestion of Pearson, decided to carry through the formula using the method of "Student." In doing this he discovered a mistake in the Tchouproff formula for the fourth power of the variance. At the same time he published [17] the formulas for the third and fourth power of the variance in the more conventional notation of that time.

It might be noted that it is particularly fitting that Church should discover this error since Tchouproff, as Pearson himself stated in an editorial [13], had pointed out a number of errors made by the Pearsonian school.

In the next volume of *Biometrika* there appears an article by Church [19] in which, among other things, formulas are derived for the third and fourth moments of the variance in the case of a finite population, sampling without replacement. Church claimed no particular credit for these formulas. His point is rather that they are almost valueless from a practical standpoint chiefly because of their length. The formula for the fourth power of the variance occupies three and one-half of the large pages of *Biometrika* and is given with the apparent aim of indicating, as Pearson said [21; 209], "the practical futility of the theoretical formulas."

Church gave full credit to Neyman for the formula for the variance of the variance and made no mention of Tchouproff's *Metron* work and of the more general presentation there given. This was particularly unfortunate because it exposed him to the charge that he ignored non-English authors. This charge was immediately made by Greenwood and Isserlis [20] who broadened it to include Neyman and, by implication, Pearson himself. They advocated the case of Tchouproff who, now dead, was unable to defend himself. They gave a survey (valuable to the cursive reader) of the pertinent contributions of the Tchouproff articles and suggested that the ignoring of Tchouproff was particularly disconcerting since it appears that Tchouproff had gone more than half way in his cooperation with English writers.

Pearson replied in an interesting article [21] which made it clear that Neyman established his results independently of Tchouproff and that the language of Neyman is much simpler than the complicated notation of Tchouproff. Pearson emphasized that Tchouproff made no attempt to give specific formulas for the third and fourth moments of the variance in the case of sampling with replacements. Pearson did not answer, at least explicitly, the claim that the Tchouproff formulas are applicable to a more general case in which there is no restriction as to the nature of the correlation between observations.

The year 1928 was marked by two important contributions. We first mention that of C. C. Craig who published his thesis in *Metron* [22]. Extending the previous results of Thiele, he was able to write the semi-invariant equivalent of the basic formulas in much less space than their previous moment formulation had demanded. He was able to write products of sample moments as well as moments of the moments themselves. His results are limited to an infinite population and one variable. The bibliography attached to his paper is commonly mentioned in later literature for its completeness. For infinite sampling it might properly be used as a supplement to the bibliography of this Part.

A most important contribution was made by R. A. Fisher [23] who was able to simplify the infinite sampling formulas greatly. He did this by introducing the sample function whose expected value is a cumulant (semi-invariant). In addition to the simplification, his ingenious attack resulted in the following contributions: (1) the recognition of the one to one correspondence between all possible independent sampling formulas and the partition of numbers, (2), that the extension of the multivariate form is accomplished by use of the partitions of multipartite numbers, (3) the tabulation of numerous new formulas, (4) the use of a general partition method by which any term in the formulas can be determined separately.

The further development of the combinatorial analysis was indicated by a paper by Fisher and Wishart which appeared in 1931 [27]. It was shown how the more involved patterns could be broken up into simpler ones.

The study of the infinite case was continued by Georgescu [28] who extended the Craig results. A feature of his work was the utilization of functions which yielded expansions of formulas in terms of successive degrees of approximation.

He applied Fisher's idea of a combinatory analysis to the conventional sample moment function.

Another paper of this series was that of Wishart [29] who gave a discriptive account of the contributions of Craig, Fisher, and Georgescu and an indication of the means of expressing the results of one writer into the language of another.

The work of Joseph Pepper which appeared in *Biometrika* in 1929 [24] should be noted. Pepper took the case of the finite parent, sampling without replacement, and two variables, and then gave an extensive list of results. He did not have a very condensed notation and was forced to assume an infinite universe for the higher moments which he studied. The important point, for historical purposes, is that Pepper combined bivariate and finite sampling. It is to be recalled that Tchouproff himself in his generalized theory gave no results for the multivariate case.

A significant advance in finite sampling was indicated by the appearance of Carver's editorial on "Fundamentals of the Theory of Sampling," which appeared in the first volume of the ANNALS OF MATHEMATICAL STATISTICS [25]. Carver took the case of a finite universe, one variable, and sampling without replacements. He presented a notation which enabled him to write the various moments of the mean through the eighth in simple form. He showed by a number of illustrations that his formula would give known results for cases both infinite and finite, when the proper restrictions were added. O'Toole [26] later generalized his results for any moment of the mean.

3. **Generalized Carver Functions and Sampling.** The use of generalized Carver functions together with the results of Part I makes possible the presentation of the general sampling theory in a compact, and yet not too symbolic, form. It is possible to write the sampling theory so that criteria 1–8 are satisfied although no attempt is made in the present paper to answer criterion 6. With reference to criteria 9–11, any affirmative answer must necessarily be tempered with qualifications as the results are far removed from that ideal solution which would permit one to determine the actual distribution of any sample moment. However the use of generalized Carver functions does permit a general concise statement of results as well as the determination of special cases. The method is also especially adapted to the introduction of new moment functions and to the use of partition analysis, although these topics are not emphasized in the present paper. In general it may be said that the use of Carver functions assists greatly in finding the theoretical sample statistics in the case of finite sampling since the Carver functions are condensed expressions of the size of the sample and the size of the parent, since they may be easily checked from symmetrical considerations, and since they are independent of the moments. They are also applicable to different replacement laws.

4. **The Use of High Moments.** Precise agreement between theoretical and practical sampling does not usually accompany the use of high moments, and

the practical statistician is apt to agree with Pearson who wrote, "I have a very firm conviction that the mathematician who uses high moments may make interesting contributions to mathematics, but he removes his work from any contact with actual statistics" [16; 117]. However since the extent of agreement between theoretical and actual results is in a sense a measure of the extent to which theoretical assumptions are actually duplicated in the experiment, it does seem sensible to discover what relations exist in the ideal theoretical case. Thiele implicitly supported the theoretical use of high moments (even in studying actual problems) when he wrote [B'; 13]:

"Therefore the general rule of the formation of good laws of presumptive errors must be:

1. In determining λ_1, and λ_2 rely almost entirely upon the actual values.

2. As to the half-invariants with high indices, say λ_4 upwards, rely as exclusively upon theoretical considerations.

3 . . . ."

A more explicit advocate is R. A. Fisher who wrote [23; 200], "In the present state of our knowledge any information, however incomplete, as to sampling distributions is likely to be of frequent use, irrespective of the fact that moment functions only provide statistical estimates of high efficiency for a special type of distribution."

Chapter II. Notation and Definition

The present chapter gives the fundamental definitions and appropriate notation. An attempt has been made to combine the most desirable features of the different notations of earlier writers.

5. Ordered Sample. An ordered sample is a sample in which distinction is made as to the order in which the variate enters the growing sample. Thus the sample found by drawing x_2 and then x_1 is the same sample as that obtained by drawing x_1 and then x_2, but it is a different ordered sample.

In some types of sampling it is possible that a given variate may appear more than once in the same sample. In general the number of ordered samples varies with the number of repeated variates. Thus the sample $x_1 + x_1$ results from but one ordered sample, while $x_1 + x_2$ results from either of two ordered samples.

6. Power Sums. Power sums have the same meaning as in section 11 of Part I. An adjustment of notation is necessary as we need to distinguish power sums of the sample from power sums of the universe. The a-th power sum of the universe is denoted by (A) while the sample power sum is denoted by (a). Similarly, bold-faced numerals are used to indicate power sums of the universe, while light-faced numerals are used to indicate power sums of the sample. The symbol $(\bar{A})$ is used to indicate that the variates are deviations from the mean of the universe.

7. Power Product Sums. Power product sums, called power products for brevity, also have the same meaning as in section 11 of Part I. Large letters are used to represent the power products of the universe while small letters are used to indicate the power products of the sample. Thus $(Q_1Q_2 \cdots Q_s)$ represents a power product of the universe while $(q_1q_2 \cdots q_s)$ represents the corresponding power product of the sample. Power products are not used extensively except in the development of the theory of the next chapter where they play an important rôle.

8. Expected Values. If a given statistical function, z is formed for every possible sample, then the arithmetic mean of the z's is the expected value of z. Thus $E(z) = \frac{\sum (z)}{S}$ where the Σ holds for all possible samples and S is the number of such samples.

9. Moments. Moments demand precise notation since distinction must be made between moments of the universe, moments of the sample, moments of the moments of the sample, and moments about the mean for these cases. In addition we wish to indicate whether or not the universe is measured about its mean.

a. *Moments of the universe.* The conventional μ's are used to indicate the moments of the universe. In this notation $\bar{\mu}$ is used to indicate the moment about the mean of the universe. Thus

$$\mu_t = \frac{(T)}{N} = \frac{\sum x^t}{N} \quad \text{and} \quad \bar{\mu}_t = \frac{(\bar{T})}{N} = \frac{\sum \bar{x}^t}{N}.$$

The usual formula relating μ_t and $\bar{\mu}_t$ [22; 20] may be written

$$\bar{\mu}_t = \sum (-1)^s \left(\frac{1^t}{(t-s)\cdot 1^s} \right) \mu_{t-s}\mu_1^s \tag{1}$$

so that

$$\bar{\mu}_2 = \frac{(2)}{N} - \frac{(1)(1)}{N^2},$$

$$\bar{\mu}_3 = \frac{(3)}{N} - \frac{3(2)(1)}{N^2} + \frac{2(1)^3}{N^3},$$

etc.

It is to be noted that, when $(1) = 0$, $\bar{\mu}_t = \mu_t$.

b. *Moments of the sample.* We denote the moments of the sample by the letter m [23; 203].

In much statistical work deviations from the mean of the universe are used in place of the variates themselves. When the universe moments about the mean appear, we indicate them with a bar. However in denoting the moments of the

samples, the moments of the mean do not appear and some other device is needed to indicate whether or not the variates are measured about the mean of the universe. The simple notations m_t and $\bar{m}_t$ are used to indicate that the variates used are deviations from the mean of the universe. A superprefix is used to indicate the case in which the variates are not measured about the mean, ${}^1m_t$, ${}^1\bar{m}_t$. The values of $\bar{m}_t$ (and ${}^1\bar{m}_t$) are obtained from the values of m_t (and ${}^1m_t$) by means of the formula

$$\bar{m}_t = \sum (-1)^s \binom{1^t}{(t-s).1^s} m_{t-s} m_1^s. \qquad \{2\}$$

c. *Moments of the moments of a sample.* Since there are many possible samples and since a given moment can be computed for each sample, it is possible to express the expected value of this moment and the expected value of any power of it. The μ's are used for this purpose. Thus

$$\begin{aligned} \mu_r(m_t) &= E(m_t)^r \\ \mu_r({}^1m_t) &= E({}^1m_t)^r \\ \mu_r(\bar{m}_t) &= E(\bar{m}_t)^r \\ \mu_r({}^1\bar{m}_t) &= E({}^1\bar{m}_t)^r. \end{aligned} \qquad \{3\}.$$

If the first one of equations $\{3\}$ represents the whole group, then the values $\bar{\mu}_r(m_t)$, $\bar{\mu}_r({}^1m_t)$, $\bar{\mu}_r(\bar{m}_t)$, and $\bar{\mu}_r({}^1\bar{m}_t)$ are indicated by

$$\bar{\mu}_r(m_t) = \sum (-1)^s \binom{1^r}{(r-s).1^s} \mu_{r-s}(m_t)\, \mu_1^s(m_t). \qquad \{4\}.$$

d. *Moments of the product of the moments of a sample.* The term $\dfrac{\sum xy}{N}$ can be indicated by $E(xy) = \mu_{11}(x, y)$. Similarly the expected value of the product of m_a and m_b may be indicated by $E(m_a m_b) = \mu_{11}(m_a, m_b)$. In general

$$\mu_{r_1 r_2 \cdots r_s}(m_{a_1}, m_{a_2}, \cdots, m_{a_s}) = E(m_{a_1}^{r_1} m_{a_2}^{r_2} \cdots m_{a_s}^{r_s}) \qquad \{5\}$$

In the case of the product of sample moment functions, when the universe is not measured about its mean, it is preferable to use a single superprefix, associated with the μ instead of a number of them associated with each m function. Thus

$$\mu_{111}({}^1m_a, {}^1m_b, {}^1m_c) = {}^1\mu_{111}(m_a, m_b, m_c).$$

The usual laws for changing from moments to moments about the mean in the case of the multivariate distributions are available. Thus

$$\bar{\mu}_{11}(m_a, m_b) = \mu_{11}(m_a, m_b) - \mu_{10}(m_a, m_b)\mu_{01}(m_a, m_b). \quad \{6\}$$

$$\begin{aligned}\bar{\mu}_{111}(m_a, m_b, m_c) = {} & \mu_{111}(m_a, m_b, m_c) - \mu_{110}(m_a, m_b, m_c)\mu_{001}(m_a, m_b, m_c) \\ & - \mu_{101}(m_a, m_b, m_c)\,\mu_{010}(m_a, m_b, m_c) \\ & - \mu_{011}(m_a, m_b, m_c)\,\mu_{100}(m_a, m_b, m_c) \\ & + 2\mu_{100}(m_a, m_b, m_c)\,\mu_{010}(m_a, m_b, m_c)\,\mu_{001}(m_a, m_b, m_c) \quad \{7\}\end{aligned}$$

etc.

10. **Different Sampling Laws.** For theoretical purposes, any law may be used in the formation of samples as long as it results in functions of all possible samples which are symmetric functions of the variates. Any uniform law of replacement satisfies this condition and hence might be used in forming samples. Most statisticians who have worked on the sampling problem have been content to assume one or the other of two replacement laws. Each of these is "natural," since it has wide application in the study of actual sampling.

The two types of sampling which have received general treatment are *sampling from an infinite universe with any law of replacement* and *sampling from a finite universe with a law of no replacements*. The results of the first type are also applicable to the case of *sampling from a finite universe when replacements are made after each drawing*. These two types of sampling have been characterized by the terms "sampling from an infinite universe," or "sampling from an unlimited supply" [25; 114] and "sampling from a finite universe" [17], or "sampling from a limited supply" [25; 101].

The theory of moments of moments for the first type of sampling has been developed to a high degree by such authors as Craig [22], Fisher [23], and Georgescu [28]. This extensive development has been due in part to the fact that the assumption of an infinite universe permits application of methods which are not applicable to the study of finite variation. The probability of getting a variate remains the same no matter what the law of replacement. The assumption of an infinite universe at first appears to make the results inapplicable to all actual problems where the universe is finite. However, if the universe is large, the assumption of infinite size does not greatly alter the results, although the extent of the change can not be determined without comparison with the results of finite sampling. A justification for the use of infinite sampling in actual finite sampling problems is based on the fact that the formulas resulting from sampling from a finite parent with replacements are the same as the infinite formulas. Hence the infinite results may be used to characterize finite sampling if *sampling is done with replacement after each drawing*. This clever scheme is somewhat invalidated, in actual sampling, because of the practicability of replacing and remixing after each drawing. Until someone demonstrates a technique which is practical and effective in securing randomness, it must be said that the value of infinite sampling theory as applied to finite

sampling depends upon the theoretically unsatisfactory assumption that a finite universe is infinite.

The theory of sampling from a finite universe without replacements has been developed by such authors as Isserlis [8], Tchouproff [15], Neyman [18], Church [19], Pepper [24], and Carver [25], although available results are not as extensive as those mentioned above because of the difficulty of algebraic manipulation and because of the length of the formulas. The fact is that the probability of getting a given variate varies with the different drawings. However, a "return to the bag" is not demanded.

The terms "infinite sampling" and "finite sampling" are adequate to describe the two kinds of sampling discussed above, but they are inadequate in the case of finite sampling if additional replacement laws are introduced. Hence, it seems preferable to characterize the type of sampling by the replacement law if the population is finite.

When the Carver functions represent known functions of n and N, it is possible to use them in writing moment formulas for any orderly replacement law. For example, it is shown in later sections how Carver functions can be applied to

1. Finite sampling without replacement,
2. Finite sampling with replacement after each drawing.
3. Finite sampling without replacement up to the n-th drawing before which the $n-1$ withdrawn variates are replaced and mixed.

The Carver function can be used symbolically even in cases in which its explicit statement in terms of n and N has not been found. In some statistical formulas the Carver functions cancel, so that the results are independent of the sampling law.

11. **Variable Distribution Laws.** It is possible to generalize the theory to include the case in which the variable takes on a different frequency distribution after each drawing, i.e., the general Tchouproff formulas can be written in terms of Carver functions. This theory can also be generalized to include many variables. In this dissertation, however, it is assumed that the universe remains the same, aside from the unreplaced variates forming the sample, throughout the sampling process.

Chapter III. The Application of the Double Expansion Theorem

It is the purpose of this chapter to establish the basic theorems on which the more specific work of the later chapters is based and to show how the double expansion theorem is to be applied to the sampling problem.

12. **Formulas Concerning Ordered Samples.** a. *Sampling with replacements.* If the samples of n are taken from a universe of N variates and if the variates are replaced after each drawing, then the number of possible ordered samples is N^n since for each of the n drawings there is a choice of N.

b. *Sampling without replacement.* If the variate is not replaced after each drawing, the number of ordered samples is

$$N(N-1)\cdots(N-n+1) = N^{(n)}.$$

c. *Replacement before the last drawing only.* In case sampling is with replacement before the last drawing only, the number of ordered samples is

$$N(N-1)\cdots(N-n+2)N = N^{(n-1)}N.$$

13. **Theorem I.** *All moments of moment functions of samples can be expressed in terms of expected values of products of power sums of samples.*

By moment functions we mean rational integral isobaric moment functions [31; 22].

The theorem follows at once from the definitions of section 9. From {3}, {4}, {5}, {6}, {7} it is clear that all moments of moment functions of samples are expressible in terms of the expected values of sample moment functions. But since the sample moment functions are themselves defined in terms of power sums of the samples, the theorem follows. For example

$$\bar{\mu}_2(\bar{m}_2) = \mu_2(\bar{m}_2) - \mu_1^2(\bar{m}_2) = E\left[\frac{(2)}{n} - \frac{(1)(1)}{n^2}\right]^2 - \left[E\left\{\frac{(2)}{n} - \frac{(1)(1)}{n^2}\right\}\right]^2 \qquad \{8\}$$

and

$$\begin{aligned}\bar{\mu}_{11}(\bar{m}_2, m_1) &= \mu_{11}(\bar{m}_2, m_1) - \mu_{10}(\bar{m}_2, m_1)\mu_{01}(\bar{m}_2, m_1)\\ &= E\left[\frac{(2)(1)}{n^2} - \frac{(1)^3}{n^3}\right] - \left[E\left\{\frac{(2)}{n} - \frac{(1)(1)}{n}\right\}\right]\left[E\frac{(1)}{n}\right] \qquad \{9\}\end{aligned}$$

14. **Theorem II.** *All moments of moment functions of samples can be expressed in terms of expected values of power products of samples.*

This follows at once from the application of the multiplication theorem of Part I to the theorem of section 13. Each product of power sums is expanded by the multiplication theorem into sums of power products. Thus

$$\begin{aligned}\mu_2('\bar{m}_2) &= E\left[\frac{(2)(2)}{n^2} - \frac{2(2)(1)(1)}{n^3} + \frac{(1)^4}{n^4}\right]\\ &= \left(\frac{1}{n^2} - \frac{2}{n^3} + \frac{1}{n^4}\right)E(4) + \left(-\frac{4}{n^3} + \frac{4}{n^4}\right)E(31) + \left(\frac{1}{n^2} - \frac{2}{n^3} + \frac{3}{n^4}\right)E(22)\\ &\qquad + \left(\frac{-2}{n^3} + \frac{6}{n^4}\right)E(211) + \frac{1}{n^4}E(1111). \qquad \{10\}\end{aligned}$$

15. **Theorem III.** *To every power product form $(q_1q_2\cdots q_s)$ there corresponds a power product form $(Q_1Q_2\cdots Q_s)$.*

The argument is simple since the terms of $(q_1q_2\cdots q_s)$ are themselves terms of $(Q_1Q_2\cdots Q_s)$. It follows at once that, if $(q_1q_2\cdots q_s)$ exists, then $(Q_1Q_2\cdots Q_s)$ exists.

As an illustration, consider the universe consisting of x_1, x_2, x_3, x_4, x_5 and the sample consisting of x_1, x_2, x_3, x_4. Then the terms of $(q_1q_2q_3) = \sum_{i_1 \neq i_2 \neq i_3}^{4} x_{i_1}^{q_1} x_{i_2}^{q_2} x_{i_3}^{q_3}$ are all contained in the terms of $(Q_1Q_2Q_3) = \sum_{i_1 \neq i_2 \neq i_3}^{5} x_{i_1}^{q_1} x_{i_2}^{q_2} x_{i_3}^{q_3}$.

16. **Theorem IV.** *If definite k's can be determined so that*

$$E(q_1q_2 \cdots q_s) = k_{p_1p_2 \cdots p_s}(Q_1Q_2 \cdots Q_s), \tag{11}$$

then it is possible to use the double expansion theorem and express the moments of the moments of the sample in terms of the P functions of Part I and the power sums (or moments) of the universe.

The double expansion theorem was designed to replace $(q_1q_2 \cdots q_s)$ by $k_{p_1p_2 \cdots p_s}(Q_1Q_2 \cdots Q_s)$. It can be used as well to replace $E(q_1q_2 \cdots q_s)$ by $k_{p_1p_2 \cdots p_s}(Q_1Q_2 \cdots Q_s)$ if the values of $k_{p_1p_2 \cdots p_s}$ can be determined. The results of such a substitution in terms of the power sums of the universe are then given by the double expansion theorem. For example

$$\mu_2({}^1m_1) = \frac{E(1)^2}{n^2} = \frac{E(2)}{n^2} + \frac{E(11)}{n^2}$$

and if $E(2) = k_1(2)$ and $E(11) = k_{11}(11)$ then

$$\mu_2({}^1m_1) = (k_2 - k_{11}) \frac{(2)}{n^2} + \frac{k_{11}(1)^2}{n^2}$$

$$= \frac{K_2}{n^2}(2) + \frac{K_{11}(1)^2}{n^2}$$

where $K_2 = k_2 - k_{11}$ and $K_{11} = k_{11}$.

It then appears that the methods and tables of Chapter I of Part I can be used in finding expressions for moments of moments, in case $k_{p_1p_2 \cdots p_s}$ is known. Thus

$$\mu_2({}^1\bar{m}_2) = E\left[\frac{(2)}{n} - \frac{(1)(1)}{n^2}\right]^2 = E\left[\frac{(2)(2)}{n^2} - \frac{2(2)(1)(1)}{n^3} + \frac{(1)(1)(1)(1)}{n^4}\right]$$

$$= \frac{P_2(4) + P_{11}(2)(2)}{n^2} - 2\left[\frac{P_3(4) + 2P_{21}(3)(1) + P_{21}(2)(2) + P_{111}(2)(1)(1)}{n^3}\right]$$

$$+ \frac{P_4(4) + 4P_{31}(3)(1) + 3P_{22}(2)(2) + 6P_{211}(2)(1)(1) + P_{1111}(1)^4}{n^4}$$

and when $(1) = 0$

$$\mu_2(\bar{m}_2) = \left(\frac{P_2}{n^2} - \frac{2P_3}{n^3} + \frac{P_4}{n^4}\right)(\bar{4}) + \left(\frac{P_{11}}{n^2} - \frac{2P_{21}}{n^3} + \frac{3P_{22}}{n^4}\right)(\bar{2})(\bar{2}). \tag{13}$$

where

$$\begin{aligned}
P_4 &= k_4 - 4k_{31} - 3k_{22} + 12k_{211} - 6k_{1111} \\
P_{31} &= k_{31} \phantom{- 3k_{22} + 1} - 3k_{211} + 2k_{1111} \\
P_{22} &= \phantom{k_4 - 4k_{31} - 3} k_{22} - 2k_{211} + k_{1111} \\
P_{211} &= \phantom{k_4 - 4k_{31} - 3k_{22} + 12} k_{211} - k_{1111} \\
P_{1111} &= \phantom{k_4 - 4k_{31} - 3k_{22} + 12k_{211} - 6} k_{1111}
\end{aligned}$$

as given by {54} of Part I.

The basic problem has thus been reduced to finding $k_{p_1 \cdots p_s}$ such that

$$E(q_1 q_2 \cdots q_s) = k_{p_1 \cdots p_s}(Q_1 Q_2 \cdots Q_s).$$

17. **Theorem V.** *The expected value of a sample power sum is always* $\frac{n}{N}$ *times the corresponding universe power sum no matter what the replacement law.*

The expected value of the sample power sum is always the same even though the k's take on different values for different replacement laws. We note first that the number of ordered samples, S, depends upon the replacement law. Now a given sample power sum, (a), has n terms, while the corresponding power sum of the universe, (A), has N terms. All the a-th powers of the variates in the universe appear in the ordered samples and, if we add all possible ordered samples, these terms appear the same number of times. Hence

$$\sum (a) = k_1^|(A) \qquad \text{and} \qquad \frac{\sum (a)}{(A)} = k_1^|.$$

Now the number of the a-th powers of the variate in $\sum (a)$ is Sn so that each of the N variates appears $\frac{Sn}{N}$ times. It follows that $\sum (a) = \frac{Sn}{N}(A)$ and hence that $E(a) = \frac{n}{N}(A)$. Hence

$$E(a) = k_1(A) \qquad \text{where} \qquad k_1 = \frac{n}{N} \tag{15}$$

no matter what the law of replacement.

An illustration may serve to clarify the argument. Consider a universe composed of x_1, x_2, x_3 and write the six ordered samples. Then

$$\frac{\sum (a)}{(A)} = \frac{x_1^a + x_2^a + x_1^a + x_3^a + x_2^a + x_1^a + x_2^a + x_3^a + x_3^a + x_1^a + x_3^a + x_2^a}{x_1^a + x_2^a + x_3^a} = 4$$

and

$$\frac{E(a)}{(A)} = \frac{2}{3} = \frac{n}{N}.$$

18. Value of $k_{p_1 \cdots p_s}$ for sampling without replacement. Consider a universe and all possible ordered samples. Form $(Q_1Q_2 \cdots Q_s)$ and $\sum (q_1q_2 \cdots q_s)$. Now $\Sigma(q_1q_2 \cdots q_s)$ is a symmetric function of the variates and consists of $N^{(n)}n^{(s)}$ products, and $(Q_1Q_2 \cdots Q_s)$ consists of $N^{(s)}$ products. Each of the $N^{(s)}$ products is repeated the same number of times in the $N^{(n)}n^{(s)}$ products of $\sum (q_1q_2 \cdots q_s)$. To find the number of times such repetition is made, it is only necessary to divide the total number of terms in $\sum (q_1q_2 \cdots q_s)$ by the number of terms in $(Q_1Q_2 \cdots Q_s)$ which gives $\frac{N^{(n)}n^{(s)}}{N^{(s)}}$. Hence

$$\sum (q_1 q_2 \cdots q_s) = \frac{N^{(n)}n^{(s)}}{N^{(s)}} (Q_1 Q_2 \cdots Q_s) \tag{16}$$

and, dividing by the number of ordered samples, $N^{(n)}$,

$$E(q_1 q_2 \cdots q_s) = \frac{n^{(s)}}{N^{(s)}} (Q_1 Q_2 \cdots Q_s) \tag{17}$$

so that

$$k_{p_1 \cdots p_s} = \frac{n^{(s)}}{N^{(s)}} \tag{18}$$

as stated in section 46 of Part I.

Since $(q_1q_2 \cdots q_s) = s_1!s_2! \cdots s_p!M(q_1q_2 \cdots q_s)$

and $(Q_1Q_2 \cdots Q_s) = s_1!s_2! \cdots s_p!M(Q_1Q_2 \cdots Q_s)$

it follows that

$$EM(q_1 q_2 \cdots q_s) = \frac{n^{(s)}}{N^{(s)}} M(Q_1 Q_2 \cdots Q_s). \tag{19}$$

Most earlier writers on finite sampling have used the idea expressed in {19} as the foundation of their work. They have found it necessary to undertake enormous algebraic manipulation to expand in terms of monomial symmetric functions and then to expand back in terms of power sums after making the coefficient adjustment. Such long derivations are not only laborious, but they are also apt to result in algebraic errors and the results obtained have not emphasized the symmetry which is inherent in the nature of the problem and which is very useful in checking calculations. It was Carver who first discovered the type of symmetric relation involved and who used it in obtaining a compact statement of the first eight moments of the sample sum in the case of a single variable. He, too, found it necessary to carry out extensive algebraic manipulations as his reference to "lavish use of symmetric functions" [25; 104] reveals. His keen insight into the essential nature of this problem led him to the conclusion that such extensive algebraic manipulation should not be necessary and that it should be possible to apply P functions to sample moments of order higher than the first. His confidence that this could be done and his

encouragement in the task have contributed in a large degree to whatever merit this dissertation may have.

With $k_{p_1\cdots p_s} = \dfrac{n^{(s)}}{N^{(s)}}$, it is at once possible to write the P function expansions. Following Carver, we let $\rho_1 = \dfrac{n}{N}$, $\rho_2 = \dfrac{n(n-1)}{N(N-1)}$, etc. and get, from sections 43 and 44 of Part I,

$$
\begin{array}{ll}
P_1 = \rho_1 & P_{11} = \rho_2 \\
P_2 = \rho_1 - \rho_2 & P_{21} = \rho_2 - \rho_3 \\
P_3 = \rho_1 - 3\rho_2 + 2\rho_3 & P_{31} = \rho_2 - 3\rho_3 + 2\rho_4 \\
P_4 = \rho_1 - 7\rho_2 + 12\rho_3 - 6\rho_4 & P_{22} = \rho_2 - 2\rho_3 + \rho_4 \\
\text{etc.} & \text{etc.}
\end{array}
$$

19. **Expected Values of Products of Sample Power Sums, Sampling Without Replacement.** The tables of Chapter I of Part I are now available for use. Thus

$$\mu_3({}^1m_1) = E({}^1m_1)^3 = \frac{1}{n^3}E(1)^3 = \frac{1}{n^3}[P_3(3) + 3P_{21}(2)(1) + P_{111}(1)^3]. \quad \{20\}$$

where

$$P_3 = \frac{n}{N} - \frac{3n(n-1)}{N(N-1)} + \frac{2n(n-1)(n-2)}{N(N-1)(N-2)}$$

$$P_{21} = \frac{n(n-1)}{N(N-1)} - \frac{n(n-1)(n-2)}{N(N-1)(N-2)}$$

$$P_{111} = \frac{n(n-1)(n-2)}{N(N-1)(N-2)}.$$

Formula {20} might be written as

$$\mu_3({}^1m_1) = \frac{1}{n^3}[P_3N\mu_3 + 3P_{21}N^2\mu_2\mu_1 + P_{111}N^3\mu_1^3] \quad \{21\}$$

We note further that as $N \to \infty$

$$NP_3 \to n, \quad P_{21}N^2 \to n(n-1), \quad P_{111}N^3 \to n(n-1)(n-2)$$

so that

$$\mu_3({}^1m_1) = \frac{1}{n^3}[n\mu_3 + 3n(n-1)\mu_2\mu_1 + n(n-1)(n-2)\mu_1^3] \quad \{22\}$$

More generally

$$P_{m_1\cdots m_r}(Q_1)(Q_2)\cdots(Q_r) = P_{m_1\cdots m_r}N^r\mu_{q_1}\mu_{q_2}\cdots\mu_{q_r}. \quad \{23\}$$

As N approaches infinity this becomes

$$P_{m_1 \cdots m_r}(Q_1)(Q_2) \cdots (Q_r) = n^{(r)} \mu_{q_1} \mu_{q_2} \cdots \mu_{q_r} . \qquad \{24\}$$

The laws of infinite sampling may be obtained by replacing power sums by moments and $P_{m_1 \cdots m_r}$ by $n^{(r)}$. The tables given in a recent paper {31; 30–32] were obtained from the tables of P functions by this method.

20. Sampling With Replacements. We next consider the case of finite sampling with replacements after each drawing. This is such a simple case that the P's can be determined without finding the k's.

Consider a universe and the N^r possible ordered samples. Thus the nine ordered samples of 2 from a universe of 3 are indicated by the subscripts

11	21	31
12	22	32
13	23	33

The samples 11, 22, 33 are not repeated while the others are. The multiplication theorem can be used in grouping types of product terms as it was in Part I, but the terms themselves have different interpretation. Thus $(1)(1) = (2) + (11)$ can be written as $(1)(1) = (2) + [11]$ where the (2) indicates the sum of the n terms found by multiplying an x by itself, while the [11] indicates the sum of the $n(n-1)$ products formed by multiplying one x by another. Since some of the x's may be alike, it is possible to have squared terms in $[1 \cdot 1]$, but they are not treated as squared terms, but rather as products. For example, if $(1) = x_1 + x_1$

$$(1)(1) = x_1^2 + x_1^2 + x_1x_1 + x_1x_1$$

so that

$$(2) = x_1^2 + x_1^2 \text{ and } [11] = x_1x_1 + x_1x_1 .$$

In determining the expected value of $(1)(1)$, we note that

$$\sum (1)(1) = \sum (2) + \sum [1 \cdot 1]$$

where $\sum$ holds for the N^n possible samples. Now $\sum (2) = k_1'(2)$ and $k_1' = \dfrac{nN^n}{N}$ so that $E(2) = \dfrac{n}{N}(2)$ as indicated in Theorem V. Also [11] is composed of $N^n n^{(2)}$ products of $\sum_{i,j=1}^{N} x_i x_j = \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{j=1}^{N} x_j\right)$. It follows that

$$\sum [1 \cdot 1] = \frac{N^n n^{(2)}}{N^2} (1)(1) \quad \text{and that}$$

$$E[1 \cdot 1] = \frac{n^{(2)}}{N^2} (1)(1).$$

It appears that $\frac{n^{(2)}}{N^2}$ plays the rôle of P_{11}.

$$\mu_2({}'m_1) = \frac{1}{n^2}E[(2) + [11]]$$

$$= \frac{1}{n^2}[P_2(2) + P_{11}(1)(1)]$$

where $P_2 = \frac{n}{N}$ and $P_{11} = \frac{n^{(2)}}{N^2}$.

The corresponding argument holds for the general case. Any product of power sums can be expanded in terms of $(q_1 q_2 \cdots q_s)$. If duplicate variates are introduced, use the notation $[q_1 q_2 \cdots q_s]$. Form $[q_1 q_2 \cdots q_s]$ for all the N^n ordered samples. Now $[q_1 q_2 \cdots q_s]$ has $n^{(s)}$ terms and $\sum [q_1 q_2 \cdots q_s] = k(Q_1)(Q_2) \cdots (Q_s)$ has $n^{(s)}N^n$ terms, while $(Q_1)(Q_2) \cdots (Q_s)$ has N^s terms. It follows that $k = \frac{n^{(s)}N^n}{N^s}$, that

$$\sum [q_1 q_2 \cdots q_s] = \frac{n^{(s)}N^n}{N^s}(Q_1)(Q_2) \cdots (Q_s),$$

and that

$$E[q_1 q_2 \cdots q_s] = \frac{n^{(s)}}{N^s}(Q_1)(Q_2) \cdots (Q_s). \qquad \{25\}$$

Hence

$$P_{m_1 \cdots m_s} = \frac{n^{(s)}}{N^s}. \qquad \{26\}$$

In general

$$P_{m_1 \cdots m_s}(Q_1)(Q_2) \cdots (Q_s) = n^{(s)}\mu_{q_1}\mu_{q_2} \cdots \mu_{q_s}. \qquad \{27\}$$

Comparison with {24} shows that the same basic laws appear no matter whether sampling is carried on with replacement, or, in the infinite case, without replacement.

21. **Other Replacement Laws.** The two cases just examined represent two extremes of orderly replacement laws. It has been shown in each case how the Carver functions can be used to express relations between the moments of the moments of the sample and the moments of the universe. It is possible to show how these functions are applicable to other replacement laws. We take, as an illustration, the case in which no replacements are made after each of the first $n - 1$ drawings, but just before the last drawing the $n - 1$ variates are replaced and mixed. I do not present here the detailed argument, but simply indicate that the appropriate value of $k_{p_1 \cdots p_s}$ is

$$k_{p_1 \cdots p_s}$$
$$= \frac{n^{(s)}}{N^{(s)}} + \frac{n-1}{N^{(s)}N}[(n-2)^{(s)} - n^{(s)} + (2^{p_1} + 2^{p_2} + \cdots + 2^{p_s})(n-2)^{(s-1)}] \qquad \{28\}$$

22. **Different Frequency Laws.** The distribution of variates may follow some known frequency law such as the normal, rectangular, binomial, Poisson, etc. In such a case, if the relations between the moments are known, it is possible to simplify the results.

Chapter IV. The Moments of the Mean

To illustrate the previous theory in a simple situation we consider the moments of the mean. Carver [25] has done this previously for the case of finite sampling without replacements, but he has taken the measures of the universe as deviations and has used the sample sum rather than the sample mean. O'Toole [26] has generalized Carver's work.

23. **The Moments of the Mean.** We have at once

$$\mu_1({}^1m_1) = \frac{1}{n} E(1) = \frac{1}{n} P_1(1) = \mu_1$$

$$\mu_2({}^1m_1) = \frac{1}{n^2} E(1)^2 = \frac{1}{n^2} [P_2(2) + P_{11}(1)(1)]$$

$$\mu_3({}^1m_1) = \frac{1}{n^3} E(1)^3 = \frac{1}{n} [P_3(3) + 3P_{21}(2)(1) + P_{111}(1)(1)(1)]$$

$$\mu_4({}^1m_1) = \frac{1}{n^4} E(1)^4 = \frac{1}{n^4} [P_4(4) + 4P_{31}(3)(1) + 3P_{22}(2)(2) + 6P_{211}(2)(1)(1) + P_{1111}(1)^4]$$

and

$$\mu_r({}^1m_1) = \frac{1}{n^r} \sum \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} P_{p_1^{\pi_1} \cdots p_s^{\pi_s}} (P_1)^{\pi_1} \cdots (P_s)^{\pi_s} \tag{29}$$

24. **Moments About the Mean of the Sample Mean.** Using {1}, we get

$$\bar{\mu}_2({}^1m_1) = \frac{1}{n^2} [P_2(2) + (P_{11} - P_1^2)(1)(1)]$$

$$\bar{\mu}_3({}^1m_1) = \frac{1}{n^3} [P_3(3) + 3(P_{21} - P_2P_1)(2)(1) + (P_{111} - 3P_{11}P_1 + 2P_1^3)(1)^3]$$

$$\bar{\mu}_4({}^1m_1) = \frac{1}{n^4} [P_4(4) + 4(P_{31} - P_3P_1)(3)(1) + 3P_{22}(2)(2) + 6(P_{211} - 2P_{21}P_1 + P_2P_1^2)(2)(1)(1) + (P_{1111} - 4P_{111}P_1 + 6P_{11}P_1^2 - 3P_1^4)(1)^4] \tag{30}$$

etc.

These formulas can be written in the notation of moments of the universe as

$$\bar{\mu}_2({}^1m_1) = \frac{1}{n^2}[P_2N\mu_2 + (P_{11} - P_1^2)N^2\mu_1^2]$$

$$\bar{\mu}_3({}^1m_1) = \frac{1}{n^3}[P_3N\mu_3 + 3(P_{21} - P_2P_1)N^2\mu_2\mu_1 + (P_{111} - 3P_{11}P_1 + 2P_1^3)N^3\mu_1^3] \quad \{31\}$$

etc.

25. Moments of the Sample Mean When the Universe is Measured About its Mean. When (1) = 0, the formulas of section {23} become

$$\mu_1(m_1) = 0$$

$$\mu_2(m_1) = \frac{1}{n^2}P_1(\bar{2})$$

$$\mu_3(m_1) = \frac{1}{n^3}P_3(\bar{3})$$

$$\mu_4(m_1) = \frac{1}{n^4}[P_4(\bar{4}) + 3P_{22}(\bar{2})^2]$$

and

$$\mu_r(m_1) = \frac{1}{n^r}\sum\binom{1^r}{p_1^{\pi_1}\ \cdots\ p_s^{\pi_s}}P_{p_1^{\pi_1}\cdots p_s^{\pi_s}}(\bar{P}_1)^{\pi_1}\cdots(\bar{P}_s)^{\pi_s} \quad \{32\}$$

where the $\sum$ holds for all partitions having no unit parts. In the language of moments {32} becomes

$$\mu_r(m_1) = \frac{1}{n^r}\sum\binom{1^r}{p_1^{\pi_1}\ \cdots\ p_s^{\pi_s}}P_{p_1^{\pi_1}\cdots p_s^{\pi_s}}N^{p_1\pi_1\cdots p_s\pi_s}(\bar{\mu}_{p_1})^{\pi_1}\cdots(\bar{\mu}_{p_s})^{\pi_s} \quad \{33\}$$

where the $\sum$ holds for all partitions of r having no unit parts.

26. Moments About the Mean of the Sample When the Universe is Measured From its Mean. Similarly, when (1) = 0, the results of section {24} become

$$\left.\begin{aligned}
\bar{\mu}_2(m_1) &= \frac{1}{n^2}P_2(\bar{2}) = \frac{1}{n^2}P_2N\bar{\mu}_2 \\
\bar{\mu}_3(m_1) &= \frac{1}{n^3}P_3(\bar{3}) = \frac{1}{n^3}P_3N\bar{\mu}_3 \\
\bar{\mu}_4(m_1) &= \frac{1}{n^4}[P_4(\bar{4}) + 3P_{22}(\bar{2})^2] \\
&= \frac{1}{n^4}[P_4N\bar{\mu}_4 + 3P_{22}N^2\bar{\mu}_2^2] \\
&\text{etc.}
\end{aligned}\right\} \quad \{34\}$$

It is to be noticed that the values $\bar{\mu}_r(m_1)$ are equal to the values $\mu_r(m_1)$. This results from (4) and the fact that $\mu_1(m_1) = 0$. It should be noted also that $\bar{\mu}_r({}^1m_1) \neq \mu_r({}^1m_1)$ as $\mu_1({}^1m_1) \neq 0$.

27. **Sampling Without Replacements.** The formulas in sections 23–26 are general formulas which become more specific as given replacement laws are introduced. If the law is sampling without replacements, we recall that $P_1 = \rho_1$, $P_2 = \rho_1 - \rho_2$, $P_3 = \rho_1 - 3\rho_2 + 2\rho_3$, etc. when $\rho_s = \dfrac{n^{(s)}}{N^{(s)}}$. It is at once possible to write the appropriate formula. Thus

$$\bar{\mu}_3(m_1) = \mu_3(m_1) = \frac{1}{n^3} P_{3,\ 3}$$

$$= \frac{1}{n^3} [\rho_1 - 3\rho_2 + 2\rho_3] N\mu_3 = \frac{(N-n)(N-2n)}{n^2(N-1)(N-2)} \bar{\mu}_3. \quad (35)$$

Now $\bar{\mu}_3 = 0$ in any symmetric universe, for example a normal or rectangular one, so $\bar{\mu}_3(m_1) = 0$.

28. **Sampling With Replacements.** In this case $P_{m_1 \cdots m_r} = \dfrac{n^{(r)}}{N^r}$ and we have

$$\mu_1({}^1m_1) = \mu_1$$

$$\mu_2({}^1m_1) = \frac{1}{n^2} [n\mu_2 + n(n-1)\mu_1^2]$$

$$\mu_3({}^1m_1) = \frac{1}{n^3} [n\mu_3 + 3n(n-1)\mu_2\mu_1 + n(n-1)(n-2)\mu_1^3]$$

$$\mu_4({}^1m_1) = \frac{1}{n^4} [n\mu_4 + 4n(n-1)\mu_3\mu_1 + 3n(n-1)\mu_2^2 + 6n(n-1)(n-2)\mu_2\mu_1^2 + n^{(4)}\mu_1^4]$$

and in general

$$\mu_r({}^1m_1) = \frac{1}{n^r} \sum \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} n^{(\rho)} (\mu_{p_1})^{\pi_1} \cdots (\mu_{p_s})^{\pi_s} \quad (36)$$

and

$$\bar{\mu}_2({}^1m_1) = \frac{1}{n^2} [n\mu_2 - n\mu_1^2]$$

$$\bar{\mu}_3({}^1m_1) = \frac{1}{n^3} [n\mu_3 - 3n\mu_2\mu_1 + 2n\mu_1^3] \quad (37)$$

$$\bar{\mu}_4({}^1m_1) = \frac{1}{n^4} [n\mu_4 - 4n\mu_3\mu_1 + 3n(n-1)\mu_2^2 - 6n(n-2)\mu_2\mu_1^2 + 3n(n-2)\mu_1^4]$$

while

$$\bar{\mu}_2(m_1) = \mu_2(m_1) = \frac{\bar{\mu}_2}{n}$$

$$\bar{\mu}_3(m_1) = \mu_3(m_1) = \frac{\bar{\mu}_3}{n^2}$$

$$\bar{\mu}_4(m_1) = \mu_4(m_1) = \frac{1}{n^3}[\bar{\mu}_4 + 3(n-1)\bar{\mu}_2^2] \tag{38}$$

etc.

29. **Sampling With Replacements Before the Last Drawing Only.** The values of $k_{p_1 \ldots p_s}$ of section 21 determine the values of the P's. Thus $P_2 = k_2 - k_{11} = \dfrac{n}{N} - \dfrac{n(n-1)}{N(N-1)} + \dfrac{2(n-1)}{N(N-1)}$ and $P_{11} = \dfrac{n(n-1)}{N(N-1)} - \dfrac{2(n-1)}{N^2(N-1)}$ so

$$\bar{\mu}_2(^1m_1) = \frac{1}{n^2}\left\{\left[n - \frac{(n-1)(n-2)}{N-1}\right]\mu_2 + \frac{(n-1)(nN-2)}{N-1}\mu_1^2\right\}. \tag{39}$$

$$\bar{\mu}_2(m_1) = \frac{1}{n^2}\left\{\left[n - \frac{(n-1)(n-2)}{N-1}\right]\bar{\mu}_2\right\}. \tag{40}$$

30. **Different Frequency Laws.** As indicated in section 22, the frequency distributions of the parent may be characterized by some moment relationship. This relationship can be inserted and the resulting formula simplified. For example, if the law of the formation of the universe is that of the hypergeometric series [25; 113]

$$\bar{\mu}_n = pq[q^{n-1} + (-1)^n p^{n-1}], \tag{41}$$

we have

$$\left.\begin{aligned} \bar{\mu}_2(m_1) &= \frac{P_2}{n^2}Npq \\ \bar{\mu}_3(m_1) &= \frac{P_3}{n^3}Npq(q^2 - p^2) \\ \bar{\mu}_4(m_1) &= \frac{1}{n^4}[P_4Npq(q^3 + p^3) + 3P_{22}N^2p^2q^2] \\ &\text{etc.} \end{aligned}\right\} \tag{42}$$

Where the values of P_2, P_3, P_4 are to be inserted according to the replacement law which is used in forming the samples. The results for sampling without replacement agree with those given by Pearson [1].

31. **Moments of the Sample Sum.** We might use the sum of the items in the sample instead of the sample mean. For example

$$\mu_2(1) = E(1)^2 = n^2E(m_1)^2 = n^2\mu_2(m_1).$$

The results would parallel the results above except that n^r in the denominator would be eliminated. It is the sample sum which is used in Carver's article [25] and this should be noted in comparing results.

Chapter V. The Mean and Variance of the Variance

As a further illustration of the use of the Carver functions there are presented in this chapter formulas for the mean of the variance and the variance of the variance.

32. The Mean of the Variance.

$$\mu_1({}^1\bar{m}_2) = E\left[\frac{(2)}{n} - \frac{(1)(1)}{n^2}\right]$$

$$= \frac{P_1}{n}(2) - \frac{P_2(2) + P_{11}(1)^2}{n^2}$$

$$= \frac{1}{n^2}[(nP_1 - P_2)(2) - P_{11}(1)^2] \quad \{43\}$$

and

$$\mu_1(\bar{m}_2) = \frac{1}{n^2}(nP_1 - P_2)N\bar{\mu}_2. \quad \{44\}$$

When sampling is with replacements $P_1 = P_2 = \frac{n}{N}$ and we get the well known

$$\mu_1(\bar{m}_2) = \frac{(n-1)}{n}\bar{\mu}_2 \quad \{45\}$$

while when sampling is without replacements, we have the well known

$$\mu_1(\bar{m}_2) = \frac{\left(1 - \frac{1}{n}\right)\bar{\mu}_2}{1 - \frac{1}{N}}. \quad \{46\}$$

33. The Second Moment of the Variance.

$$\mu_2({}^1\bar{m}_2) = E\left[\frac{(2)^2}{n^2} - \frac{2(2)(1)(1)}{n^3} + \frac{(1)^4}{n^4}\right]$$

becomes

$$\mu_2({}^1\bar{m}_2) = \left(\frac{P_2}{n^2} - \frac{2P_3}{n^3} + \frac{P_4}{n^4}\right)(4) - 4\left(\frac{P_{21}}{n^3} - \frac{P_{31}}{n^4}\right)(3)(1)$$

$$+ \left(\frac{P_{11}}{n^2} - \frac{2P_{21}}{n^3} + \frac{3P_{22}}{n^4}\right)(2)(2) - 2\left(\frac{P_{111}}{n^3} - \frac{3P_{211}}{n^4}\right)(2)(1)(1) + P_{1111}(1)^4 \quad \{47\}$$

and

$$\mu_2(\bar{m}_2) = \left(\frac{P_2}{n^2} - \frac{2P_3}{n^3} + \frac{P_4}{n^4}\right)(\bar{\mathbf{4}}) + \left(\frac{P_{11}}{n^2} - \frac{2P_{21}}{n^3} + \frac{3P_{22}}{n^4}\right)(\bar{\mathbf{2}})^2. \quad \{48\}$$

These of course can be written in terms of moments of the variance.

34. The Variance of the Variance. Since $\bar{\mu}_2({}^1\bar{m}_2) = \mu_2({}^1\bar{m}_2) - \mu_1^2({}^1\bar{m}_2)$, we have

$$\begin{aligned}\bar{\mu}_2({}^1\bar{m}_2) &= \left(\frac{P_2}{n^2} - \frac{2P_3}{n^3} + \frac{P_4}{n^4}\right)(\mathbf{4}) - 4\left(\frac{P_{21}}{n^3} - \frac{P_{31}}{n^4}\right)(\mathbf{3})(\mathbf{1}) \\ &\quad + \left[\frac{P_{11}}{n^2} - \frac{2P_{21}}{n^3} + \frac{3P_{22}}{n^4} - \left(\frac{P_1}{n} - \frac{P_2}{n^2}\right)^2\right](\mathbf{2})(\mathbf{2}) \\ &\quad - 2\left[\frac{P_{111}}{n^3} - \frac{3P_{211}}{n^4} - \frac{P_1P_{21}}{n^3} + \frac{P_2P_{11}}{n^4}\right)(\mathbf{2})(\mathbf{1})(\mathbf{1}) + \frac{P_{1111} - P_{11}^2}{n^4}(\mathbf{1})^4. \quad \{49\}\end{aligned}$$

Formula {49} may also be written as

$$\begin{aligned}\bar{\mu}_2({}^1\bar{m}_2) &= \frac{1}{n^4}\{(n^2P_2 - 2nP_3 + P_4)N\mu_4 - 4(nP_{21} - P_{31})N^2\mu_3\mu_1 \\ &\quad + (n^2P_{11} - 2nP_{21} + 3P_{22} - n^2P_1^2 + 2nP_1P_2 - P_2^2)N^2\mu_2^2 \\ &\quad - 2(nP_{111} - 3P_{211} - nP_1P_{11} + P_2P_{11})N^3\mu_2\mu_1^2 + (P_{1111} - P_{11}^2)N^4\mu_1^4\}. \quad \{50\}\end{aligned}$$

Formulas {49} and {50} are not expressed in terms of deviations of the variates. Neither do they assume any particular replacement law nor any particular type of universe.

In case the universe is measured about its mean we can write at once, by placing $(\mathbf{1}) = 0$ in {49}

$$\begin{aligned}\bar{\mu}_2(\bar{m}_2) &= \left(\frac{P_2}{n^2} - \frac{2P_3}{n^3} + \frac{P_4}{n^4}\right)(\bar{\mathbf{4}}) \\ &\quad + \left[\left(\frac{P_{11}}{n^2} - \frac{2P_{21}}{n^3} + \frac{3P_{22}}{n^4}\right) - \left(\frac{P_1}{n} - \frac{P_2}{n^2}\right)^2\right](\bar{\mathbf{2}})(\bar{\mathbf{2}}) \quad \{51\}\end{aligned}$$

and

$$\begin{aligned}\bar{\mu}_2(\bar{m}_2) &= \frac{1}{n^4}\{(n^2P_2 - 2nP_3 + P_4)N\bar{\mu}_4 + (n^2P_{11} - 2nP_{21} + 3P_{22} - n^2P_1^2 \\ &\quad + 2nP_1P_2 - P_2^2)N\bar{\mu}_2^2\}. \quad \{52\}\end{aligned}$$

35. Sampling Without Replacements. Using the P's as defined by sampling without replacements, it appears that the coefficient of the μ_4 term

$$\left(\frac{P_2}{n^2} - \frac{2P_3}{n^3} + \frac{P_4}{n^4}\right)N = \frac{N}{n^3}\frac{(N-n)(n-1)(Nn - N - n - 1)}{(N-1)(N-2)(N-3)} \quad \{53\}$$

agrees with that given by Neyman [18; 477], Tchouproff [15; 660], Pepper [24; 234], Carver [25; 270]. Also the coefficient of the μ_2^2 term

$$\frac{P_{11}}{n^2} - \frac{2P_{21}}{n^3} + \frac{3P_{22}}{n^4} - \left(\frac{P_1}{n} - \frac{P_2}{n^2}\right)^2 N^2$$
$$= -\frac{N(N-n)(n-1)(N^2n - 3N^2 + 6N - 3n - 3)}{n^3(N-1)^2(N-2)(N-3)} \qquad \{54\}$$

agrees with that of the above authors.

As far as the author is aware, no one has written the coefficients of $\mu_3\mu_1$, $\mu_2\mu_1^2$, and μ_1^4 in the formula for $\bar{\mu}_2('\bar{m}_2)$.

The coefficient of $\mu_3\mu_1$ is

$$-4\left(\frac{P_{21}}{n^3} - \frac{P_{31}}{n^4}\right)N^2 = \frac{-4N(n-1)(N-n)(Nn-N-n-1)}{n^3(N-1)(N-2)(N-3)}. \qquad \{55\}$$

The coefficient of $\mu_2\mu_1^2$ is

$$-2\left(\frac{P_{111}}{n^3} - \frac{3P_{211}}{n^4} - \frac{P_1P_{11}}{n^3} + \frac{P_2P_{11}}{n^4}\right)N^3$$
$$= \frac{4N^2}{n^3}\,\frac{(n-1)(N-n)[(2n-3)N-3(n-1)]}{(N-1)^2(N-2)(N-3)} \qquad \{56\}$$

while the coefficient of μ_1^4 is

$$\left(\frac{P_{1111}}{n^4} - \frac{P_{11}P_{11}}{n^4}\right)N^4 = -\frac{2N^2}{n^3}\,\frac{(n-1)(N-n)[(2n-3)N-3(n-1)]}{(N-1)^2(N-2)(N-3)}. \qquad \{57\}$$

It is possible with some algebraic manipulation to use the P functions to express the coefficients of the moments as functions of N and n. The suggestion here is that such algebraic work is unnecessary since the left members of {53} $\cdots$ {57} are as easily handled in an actual problem as the right hand members. It is possible to compute the coefficients from the ρ's and the P's without writing explicit expansions in terms of N and n. Besides the formulas involving N and n are so lengthy that algebraic errors are apt to occur. The use of Carver functions is further advocated because the same basic formulas are applicable to all types of sampling and because the tables of Chapter I of Part I are directly applicable.

36. Sampling With Replacements. If $P_{m_1\cdots m_r} = \frac{n^{(r)}}{N^{(r)}}$, the coefficient of μ_4 is $\frac{1}{n^4}[(n(n-1)^2] = \frac{(n-1)^2}{n^3}$ while the coefficient of μ_2^2 is

$$\frac{1}{n^4}[(n^2-2n+3)n(n-1) - (n^2-2n+1)n^2] = \frac{(n-1)(3-n)}{n^3}.$$

Then {52} becomes

$$\bar{\mu}_2(\bar{m}_2) = \frac{1}{n^3}[(n-1)^2\bar{\mu}_4 - (n-1)(n-3)\bar{\mu}_2^2]. \qquad \{58\}$$

The formula for $\bar{\mu}_2({}^{|}\bar{m}_2)$ becomes

$$\bar{\mu}_2({}^{|}\bar{m}_2) = \frac{1}{n^3}[(n-1)^2\mu_4 - 4(n-1)^2\mu_3\mu_1 - (n-1)(n-3)\mu_2^2$$
$$+ 4(2n-3)(n-1)\mu_2\mu_1^2 - 2(2n-3)(n-1)\mu_1^4]. \qquad \{59\}$$

Now {58} can be written in terms of semi-invariants by the use of $\bar{\mu}_4 = \lambda_4 + 3\lambda_2^2$ and $\bar{\mu}_2 = \lambda_2$ so

$$\bar{\mu}_2(\bar{m}_2) = \frac{1}{n^3}[(n-1)^2\lambda_4 + 2n(n-1)\lambda_2^2].$$

See [B′; 209], [22; 57].

37. Different Distribution Laws. Given frequency laws can be inserted. Thus {44} becomes

$$\mu_1(\bar{m}_2) = \frac{1}{n^2}(nP_1 - P_2)pq \qquad \text{if the} \qquad \bar{\mu}_2 = pq$$

while {52} becomes, if $\bar{\mu}_2 = pq$ and $\bar{\mu}_4 = pq(q^3 + p^3)$

$$\bar{\mu}_2(\bar{m}_2) = \frac{N}{n^4}(n^2P_2 - 2nP_3 + P_4)pq(q^3 + p^3)$$
$$+ \frac{N^2}{n^4}(n^2P_{11} - 2nP_{21} + 3P_{22} - n^2P_1^2 + 2nP_2P_1 - P_2^2)p^2q^2. \qquad \{60\}$$

Other frequency laws can be inserted similarly.

Chapter VI. Tabular Presentation of Formulas.

It is the purpose of this dissertation to show how the P functions can be used in finite sampling rather than to present an exhaustive list of formulas. The specific formulas of the two previous chapters are derived, primarily, for illustrative purposes. The implication is that other formulas may be derived similarly.

However, it is possible to present, implicitly in tabular form, a number of formulas. In this chapter there are presented formulas involving moments of weight equal to or less than 6.

38. The formulas of weight 2.

$$\left.\begin{aligned} \mu_1({}^{|}\bar{m}_2) &= \left(\frac{P_1}{n} - \frac{P_2}{n^2}\right)(2) - \frac{P_{11}}{n^2}(1)(1) \\ \mu_2({}^{|}m_1) &= \left[\frac{P_2}{n^2}(2) + \frac{P_{11}}{n^2}(1)^2\right] \end{aligned}\right\} \qquad \{61\}$$

can be written in tabular form as

	2	11		2	11
2	P_1			$\frac{1}{n}$	
11	P_2	P_{11}		$-\frac{1}{n^2}$	$\frac{1}{n^2}$

with little effort. The first entries in the top row indicate the power sums of the universe, while the columnar entries indicate the moments of the sample. Now

$$\mu_1({}^1\bar{m}_2) = E\left[\frac{(2)}{n} - \frac{(1)(1)}{n^2}\right]$$

and

$$\mu_2(m_1) = E\left[\frac{(1)(1)}{n^2}\right].$$

The coefficients of the power sums in the expansion of $\bar{m}$ are entered in the right hand part of the table. Thus, under 2, there appear the entries $\frac{1}{n}$ and $-\frac{1}{n^2}$. These when multiplied by the power sums as indicated on the left, give $\bar{m}_2 = \frac{(2)}{n^2} - \frac{(1)(1)}{n^2}$. Similarly $m_1^2 = \frac{(1)(1)}{n^2}$.

Now the expected value is given by the proper P function expansion. The left hand portion of the table, which is the same as the P function table of Chapter I of Part I, gives such expansions. Thus the coefficient of (2) in $E(m_2)$ is $\frac{P_1}{n} - \frac{P_2}{n^2}$, while the coefficient of (1)(1) is $-\frac{P_{11}}{n^2}$. Hence the complete formula is

$$\mu_1({}^1\bar{m}_2) = \left(\frac{P_1}{n} - \frac{P_2}{n^2}\right)(2) - \frac{P_{11}}{n^2}(1)(1)$$

as indicated above.

39. The Formulas of Weight 3. Similarly the table

	3	21	111		3	21	111
3	P_1				$\frac{1}{n}$		
21	P_2	P_{11}			$-\frac{3}{n^2}$	$\frac{1}{n^2}$	
111	P_3	$3P_{21}$	P_{111}		$\frac{2}{n^3}$	$-\frac{1}{n^3}$	$\frac{1}{n^3}$

can be used to give the formulas

$$\mu_1({}^{|}\bar{m}_3) = \left(\frac{P_1}{n} - \frac{3P_2}{n^2} + \frac{2P_3}{n^3}\right)(3) - 3\left(\frac{P_{11}}{n^2} - \frac{2P_{21}}{n^3}\right)(2)(1) + 2\frac{P_{111}}{n^3}(1)^3 \quad \{62\}$$

$${}^{|}\mu_{11}(\bar{m}_2, m_1) = \left(\frac{P_2}{n^2} - \frac{P_3}{n^3}\right)(3) + \left(\frac{P_{11}}{n^2} - \frac{3P_{21}}{n^3}\right)(2)(1) - \frac{2P_{111}}{n^3}(1)^3 \quad \{63\}$$

$$\mu_3({}^{|}m_1) = \frac{P_3}{n^3}(3) + \frac{3P_{21}}{n^3}(2)(1) + \frac{P_{111}}{n^3}(1)^3. \quad \{64\}$$

In case we wish to express the results in terms of moments about the mean, (1) = 0, and we have

$$\mu_1(\bar{m}_3) = \left(\frac{P_1}{n} - \frac{3P_2}{n^2} + \frac{2P_3}{n^3}\right)(\bar{3}) \quad \{65\}$$

$$\mu_{11}(\bar{m}_2, m_1) = \left(\frac{P_2}{n^2} - \frac{P_3}{n^3}\right)(\bar{3}) \quad \{66\}$$

$$\mu_3(m_1) = \frac{P_3}{n^3}(\bar{3}) \quad \{67\}$$

so that

$$\mu_1(\bar{m}_3) = \left(\frac{P_1}{n} - \frac{3P_2}{n^2} + \frac{2P_3}{n^3}\right)N\bar{\mu}_3 \quad \{68\}$$

$$\mu_{11}(\bar{m}_2, m_1) = \left(\frac{P_2}{n^2} - \frac{P_3}{n^3}\right)N\bar{\mu}_3 \quad \{69\}$$

$$\mu_3(m_1) = \frac{P_3}{n^3}N\bar{\mu}_3. \quad \{70\}$$

The insertion of specific sampling laws gives the specific results of earlier authors.

40. **The Tabular Forms.** It is further evident that the power of n in the denominator is equal to the sum of the subscripts of the Carver function above it. We might utilize this knowledge and write in the right hand part of the table the numerators of the entries in the tables above. The table of weight of 3 would then appear as

	3	21	111		3	21	111
3	P_1				1		
21	P_2	P_{11}			-3	1	
111	P_3	$3P_{21}$	P_{111}		2	-1	1

and it is possible to read {62}, {63}, {64}, {65}, {66}, and {67} directly from it.

TABLE I

$w = 2$

	2	11		2	11
2	P_1			1	
11	P_2	P_{11}		-1	1

$w = 3$

	3	21	1		3	21	1
3	P_1				1		
21	P_2	P_{11}			-3	1	
1	P_3	$3P_{21}$	P_{111}		2	-1	1

$w = 4$

	4	31	22	211	1111		4	31	22	211	1111
4	P_1						1				
31	P_2	P_{11}					-4	1			
22	P_2		P_{11}						1		
211	P_3	$2P_{21}$	P_{21}	P_{111}			6	-3	-2	1	
1111	P_4	$4P_{31}$	$3P_{22}$	$6P_{211}$	P_{1111}		-3	2	1	-1	1

$w = 5$

	5	41	32	311	221	2111	1^5		5	41	32	311	221	2111	1^5
5	P_1								1						
41	P_2	P_{11}							-5	1					
32	P_2		P_{11}								1				
311	P_3	$2P_{21}$	P_{21}	P_{111}					10	-4	-1	1			
221	P_3	P_{21}	$2P_{21}$		P_{111}						-3		1		
2111	P_4	$3P_{31}$	$P_{31} + 3P_{22}$	$3P_{211}$	$3P_{211}$	P_{1111}			-10	6	5	-3	-2	1	
11111	P_5	$5P_{41}$	$10P_{32}$	$10P_{311}$	$15P_{221}$	$10P_{2111}$	P_{11111}		4	-3	-2	2	1	1	1

$w = 6$

	6	51	42	33	411	321	222	3111	2^21^2	$2\,1^4$	1^6		6	51	42	33	411	321	222	3111	2^21^2	$2\,1^4$	1^6
6	P_1												1										
51	P_2	P_{11}											-6	1									
42	P_2		P_{11}												1								
33	P_2			P_{11}												1							
411	P_3	$2P_{21}$	P_{21}		P_{111}								15	-5	-1		1						
321	P_3	P_{21}	P_{21}	P_{21}		P_{111}									-4	-6		1					
222	P_3		$3P_{21}$				P_{111}												1				
3111	P_4	$3P_{31}$	$3P_{22}$	P_{31}	$3P_{211}$	$3P_{211}$		P_{1111}					20	10	4	4	-4	-1		1			
2211	P_4	$2P_{31}$	$P_{22} + 2P_{31}$	$2P_{22}$	P_{211}	$4P_{211}$	P_{211}		P_{1111}						6	9		-3	-3		1		
21^4	P_5	$4P_{41}$	$P_{41} + 6P_{32}$	$4P_{32}$	$6P_{311}$	$4P_{311} + 12P_{221}$	$3P_{221}$	$4P_{2111}$	$6P_{2111}$	P_{11111}			-15	-10	-9	-12	6	5	3	-3	-2	1	
1^6	P_6	$6P_{51}$	$15P_{42}$	$10P_{33}$	$15P_{411}$	$60P_{321}$	$15P_{222}$	$20P_{3111}$	$45P_{2211}$	$15P_{21111}$	P_{111111}		5	4	3	4	-3	-2	-1	2	1	-1	1

The tables of weight $W = 2, 3, 4, 5, 6$ are given in Table I. The right hand partitions not involving unit parts are underscored as these indicate the columns which should be used if the universe is measured about its mean. As an illustration we write from Table I the value of $\mu_2(\bar{m}_2)$. We get

$$\mu_2(\bar{m}_2) = \left(\frac{P_2}{n^2} - \frac{2P_3}{n^3} + \frac{P_4}{n^4}\right) N\bar{\mu}_4 + \left(\frac{P_{11}}{n^2} - \frac{2P_{21}}{n^3} + \frac{3P_{22}}{n^4}\right) N\bar{\mu}_2^2$$

as previously indicated.

The same tabular scheme can be used to write formulas of weight greater than 6.

41. **Moments of Other Sample Moment Functions.** It is possible to use a similar tabular scheme when we wish to find the moments of other sample moment functions. We define

$$l_2 = \frac{(2)}{n} - \frac{(1)(1)}{n^2}$$

$$l_3 = \frac{(3)}{n} - \frac{3(2)(1)}{n^2} + \frac{2(1)^3}{n^3}$$

$$l_4 = \frac{(4)}{n} - \frac{4(3)(1)}{n^2} - \frac{3(2)(2)}{n^2} + \frac{12(2)(1)(1)}{n^3} - \frac{6(1)^4}{n^4}$$

and, in general,

$$l_r = \sum (-1)^{\rho-1}(\rho - 1)! \binom{1^r}{p_1^{\pi_1} \cdots p_s^{\pi_s}} \frac{(p_1)^{\pi_1} \cdots (p_s)^{\pi_s}}{n^\rho}. \tag{71}$$

The formulas of weight 5 are given by Table II.

TABLE II

	5	41	32	311	221	2111	1^5		−5	41	32	311	221	21^3	1^5
5	P_1								1						
41	P_2	P_{11}							−5	1					
32	P_2		P_{11}						−10		1				
311	P_3	$2P_{21}$	P_{21}	P_{111}					20	−4	−1	1			
221	P_3	P_{21}	$2P_{21}$		P_{111}				30	−3	−3		1		
2111	P_4	$3P_{31}$	$P_{31} + 3P_{22}$	$3P_{211}$	$3P_{211}$	P_{1111}			−60	12	5	−3	−2	1	
11111	P_5	$5P_{41}$	$10P_{32}$	$10P_{311}$	$15P_{221}$	$10P_{2111}$	P_{11111}		24	−6	−2	2	1	−1	1

Thus for example

$$
{}^{1}\mu_{11}(l_4, l_1) = \left(\frac{P_2}{n^2} - \frac{7P_3}{n^3} + \frac{12P_4}{n^4} - \frac{6P_5}{n^5}\right) N\bar{\mu}_5
$$

$$
+ \left(-\frac{10P_{21}}{n^3} + \frac{12P_{31}}{n^4} + \frac{36P_{22}}{n^4} - \frac{60P_{32}}{n^5}\right) N^2 \bar{\mu}_3 \bar{\mu}_2. \quad \{72\}
$$

If all the entries in the right hand part of Table I, except the unit terms in the main diagonal, are placed equal to 0, the tables can be used to give the moment function of the m_a. Thus, when $w = 3$,

$$
\mu_1({}^1m_3) = \frac{P_1}{n}(3) \qquad \{73\}
$$

$$
{}^1\mu_{11}(m_2, m_1) = \frac{P_2}{n^2}(3) + \frac{P_{11}}{n^2}(2)(1) \qquad \{74\}
$$

$$
\mu_3({}^1m_1) = \frac{P_3}{n^3}(3) + \frac{3P_{21}}{n^3}(2)(1) + \frac{P_{111}}{n^3}(1)^3 \qquad \{75\}
$$

and

$$
\mu_1(m_3) = \frac{P_1}{n} N\bar{\mu}_3 \qquad \{76\}
$$

$$
\mu_{11}(m_2, m_1) = \frac{P_2}{n^2} N\bar{\mu}_3 \qquad \{77\}
$$

$$
\mu_3(m_1) = \frac{P_3}{n^3} N\bar{\mu}_3. \qquad \{78\}
$$

42. Other Moment Functions. The tables give such formulas as $\mu_r(\bar{m}_a)$, $\mu_{r_1r_2}(\bar{m}_a, \bar{m}_b)$, etc. If formulas for $\bar{\mu}_r(\bar{m}_a)$, $\bar{\mu}_{r_1r_2}(\bar{m}_a, \bar{m}_b)$ etc., are needed, it is necessary to go through the usual work of changing from moments to moments about the mean.

Let us derive a general formula for the correlation of the mean and the variance as an illustration of the use of the tabular formulas. By definition

$$
r_{11}(\bar{m}_2, m_1) = \frac{\bar{\mu}_{11}(\bar{m}_2, m_1)}{[\bar{\mu}_{20}(\bar{m}_2, m_1)\bar{\mu}_{02}(\bar{m}_2, m_1)]^{\frac{1}{2}}}. \qquad \{79\}
$$

Now

$$
\begin{aligned}
\bar{\mu}_{11}(\bar{m}_2, m_1) &= \mu_{11}(\bar{m}_2, m_1) \\
\bar{\mu}_{20}(\bar{m}_2, m_1) &= \mu_2(\bar{m}_2) - \mu_1^2(\bar{m}_2) \\
\bar{\mu}_{02}(\bar{m}_2, m_1) &= \mu_2(m_1) - \mu_1^2(m_1) = \mu_2(m_1).
\end{aligned}
$$

Some of these values have appeared earlier in this paper. Without using the earlier results, we find from Table I

$$\mu_{11}(\bar{m}_2, m_1) = \left(\frac{P_2}{n^2} - \frac{P_3}{n^3}\right) N\bar{\mu}_3$$

$$\mu_2(\bar{m}_2) = \left(\frac{P_2}{n^2} - \frac{2P_3}{n^3} + \frac{P_4}{n^4}\right) N\bar{\mu}_4 + \left(\frac{P_{11}}{n^2} - \frac{2P_{21}}{n^3} + \frac{3P_{22}}{n^4}\right) N^2 \bar{\mu}_2^2$$

$$\mu_1(\bar{m}_2) = \left(\frac{P_1}{n} - \frac{P_2}{n^2}\right) N\bar{\mu}_2$$

$$\mu_2(m_1) = \frac{P_2}{n^2} N\bar{\mu}_2.$$

Hence {79} becomes

$$r_{11}(\bar{m}_2, m_1) = \frac{(nP_2 - P_3)\bar{\mu}_3}{[(n^2P_2^2 - 2nP_2P_3 + P_2P_4)\bar{\mu}_4\bar{\mu}_2 - (n^2P_2P_{11} - 2nP_2P_{21} + 3P_2P_{22} - n^2P_2P_1^2 + 2nP_2^2P_1 - P_2^3)N\bar{\mu}_2^3]^{\frac{1}{2}}} \quad \{80\}$$

Formula {80} gives the correlation between the variance and the mean no matter what the law of replacement. If the universe is symmetric, $\bar{\mu}_3 = 0$ and $r_{11}(\bar{m}_2, m_1) = 0$.

The usual special cases may be obtained. When replacements are made, {80} becomes at once

$$r_{11}(\bar{m}_2, m_1) = \frac{(n-1)\bar{\mu}_3}{[(n-1)\bar{\mu}_4\bar{\mu}_2 - (3-n)\bar{\mu}_2^3]^{\frac{1}{2}}} \quad \{81\}$$

as indicated by Pepper [24; 246].

When no replacements are made {80} reduces to results previously given by Neyman [18; 489] and Pepper [24; 245].

43. Conclusion. The theory presented here is capable of generalization in many ways. For example, application to multivariate distributions readily follows. However an attempt has been made in this dissertation to emphasize the essence of the method. Illustrations have been chosen to indicate its inherent generality.

It should be stated, finally, that the aim of this dissertation is not primarily to provide a list of sampling formulas, but rather to provide a method by which the desired sampling formula may be derived without too much algebraic work.

In concluding this dissertation, I wish to acknowledge the guidance and encouragement of Professsor H. C. Carver. Also I wish to express my appreciation to Professor R. A. Fisher and to Professor C. C. Craig, who read the manuscript, or portions of it, and made needed suggestions for improvement. I am also indebted to Professor J. A. Nyswander and Professor T. H. Hildebrandt for valuable advice and assistance.

BIBLIOGRAPHY

(1) Pearson, K., "On Certain Properties of the Hypergeometric Series etc., *Philosophical Magazine*, 45 (1899), pp. 231–246.

(2) Pearson, K., "On the Probable Error of Frequency Constants," *Biometrika*, 2 (1902–3), pp. 273–281.

(3) Henderson, R., "Frequency Curves and Moments," *Transactions of Acturial Society of America*," 8 (1904), pp. 30–42.

(4) Pearson, K., "Note on the Significant or Non-significant Character of a Sub-sample Drawn from a Sample." *Biometrika*, 5 (1906), pp. 181–183.

(5) "Student," "The Probable Error of a Mean," *Biometrika*, 6 (1908) pp. 1–25.

(6) Pearson, K., "On the Probable Errors of Frequency Constants." *Biometrika*, 9 (1913), pp. 1–10.

(7) Isserlis, L., On the Conditions Under Which the Probable Error of Frequency Distributions Have Real Significance." *Royal Society Proceedings*, 92A (1915), pp. 23–41.

(8) Isserlis, L., "On the Value of a Mean as Calculated From a Sample," *Journal of Royal Statistical Society*, 81 (1918), pp. 75–81.

(9) Edgeworth, F., "On the Value of a Mean as Calculated from a Sample," *Journal of Royal Statistical Society*, 81 (1918), pp. 624–632.

(10) Tchouproff, A., "On the Mathematical Expectation of a Positive Integral Power of the Difference Between the Frequency and the Probability of an Event." *Proceedings of the Petrograd Polytechnic Institute*,

(11) Tchouproff, A., "Zur Theorie der Stabilitat Statisticher Reirhen," *Skandinavisk Aktuarietidskrift*, (1918).

(12) Tchouproff, A., "On the Mathematical Expectation of the Moments of Frequency Distributions." Part I *Biometrika*, 12 (1918), pp. 140–169, 185–210. Part II, *Biometrika* 13 (1920–21), pp. 283–295.

(13) Pearson, K., "Peccavimus." Biometrika, 12 (1918–19), pp. 259–281.

(14) Splawa-Neyman, J., "*La Revue Mensuelle de Statistique.*" Tome 6 (1923) pp. 1–29.

(15) Tchouproff, A., "On the Mathematical Expectation of the Moments of Frequency Distributions in the Case of Correlated Observations," Metron 2 (1923), pp. 461–493; 646–683.

(16) Pearson, K., "Note on Professor Romanovsky's Generalization of my Frequency Curves," *Biometrika*, 16 (1924), pp. 116–117.

(17) Church, A., "On the Moments of the Distributions of Squared Deviations for Samples of N Drawn from an Indefinitely Large Population," *Biometrika*, 17 (1925), pp. 79–83.

(18) Neyman, J., "Contributions to the Theory of Small Samples Drawn From a Finite Population," *Biometrika*, 17 (1925), pp. 472–479.

(19) Church, A., "On the Means and Squared Deviations of Small Samples From Any Population," *Biometrika*, 18 (1926), pp. 321–394.

(20) Greenwood, M., and Isserlis, L., "A Historical Note on the Problems of Small Samples," *Royal Statistical Society Journal*, 90 (1927), pp. 347–352.

(21) Pearson, K., "Another Historical Note on the Problem of Small Samples," *Biometrika*, 19 (1927), pp. 207–210.

(22) Craig, C. C., "An Application of Thiele's Semi-invariants to the Sampling Problem," *Metron*, 7 (1928–29), pp. 3–74.

(23) Fisher, R. A., "Moments and Product Moments of Sampling Distributions," *Proceedings London Mathematical Society*, 2 (30) (1929), pp. 199–238.

(24) Pepper, J., "Studies in the Theory of Sampling," *Biometrika*, 21 (1929), pp. 231–258.

(25) Carver, H. C., "Fundamentals of the Theory of Sampling," *Annals of Mathematical Statistics*, 1 (1930), pp. 101–121; 260–274.

(26) O'Toole, A. L., "On Symmetric Functions and Symmetric Functions of Symmetric Functions," *Annals of Mathematical Statistics*, 2 (1931), pp. 102–149.
(27) Fisher, R. A., and Wishart, J., "The Derivation of the Pattern Formulae of Two Way Partitions From Those of Simpler Patterns," *Proceedings London Mathematical Society*, 2–33 (1931), pp. 195–208.
(28) St. Georgescu, N., "Further Contributions to the Sampling Problem." *Biometrika*, 24 (1932), pp. 65–107.
(29) Wishart, J., "A Comparison of the Semi-invariants of the Distributions of the Moment and the Semi-invariant Estimates in Sampling From an Infinite Population," *Biometrika*, 25 (1933), pp. 52–60.
(30) Feldman, H., "Mathematical Expectations of Product Moments of Samples Drawn from a Set of Infinite Populations." *Annals of Mathematical Statistics*, 6 (1935), pp. 30–52.
(31) Dwyer, P. S., "Moments of Any Rational Integral Isobaric Sample Moment Function," *Annals of Mathematical Statistics*, 8 (1937), pp. 21–65.

BOOKS

A. Whitworth, "Choice and Chance," 4th Edition (1886).
B. Thiele, T. N., "Theory of Observations," (1903).
B'. Reprinted *Annals of Mathematical Statistics*, 2 (1931), pp. 165–306.
C. Mortara, "Elementi di Statistica," Roma (1917).
D. Rietz, H. L., (Editor-in-chief), "Handbook of Mathematical Statistics" 1924.
E. Rietz, H. L., "Mathematical Statistics," (1927).

DISTRIBUTIONS OF SUMS OF SQUARES OF RANK DIFFERENCES FOR SMALL NUMBERS OF INDIVIDUALS[1]

By E. G. Olds

I. INTRODUCTION

In a recent article,[2] reporting the results of research under a grant-in-aid from the Carnegie Corporation of New York, Hotelling and Pabst have given a comprehensive treatment of the theory and application of rank correlation and have contributed significantly to existing knowledge on the subject. It is not the purpose of this note to evaluate their contribution but to attempt the solution of a problem they suggest.

In §3[3] they have given the well-known formula for rank correlation, $r' = 1 - \frac{6\Sigma d^2}{n^3 - n}$, where n is the number of individuals ranked and $\Sigma d^2 = \sum_{i=1}^{n} d_i^2$ (d_i being the rank difference for the ith individual). In §5 the question of the significance of r' in small samples has been considered from the following point of view; if the value of r', obtained from a comparison of the ranks of n individuals as a possible measure of the relation between two attributes, is such that there exists a high probability that it could have occurred by virtue of a chance rearrangement of the n individuals, then the value of r' does not furnish a significant indication of relationship. Then one test of the significance of a particular value of r' is to note whether it has a probability less than P (P equal to .01 or, less stringently, equal to .05) of occurring because of a chance re-ranking.

To apply this test it is necessary to have some information regarding the distribution of r' for the chance rearrangements of the numbers from 1 to n. Hotelling and Pabst have given the distribution of r' for the cases, $n = 2, 3, 4$. They have noted that the distribution is symmetrical for each value of n and that it has a range from -1 to 1. From a consideration of the probabilities corresponding to $\Sigma d^2 = 0, 2, 4, 6$, they have discussed the significance of values of r' for $n = 5, 6, 7$. In §8 they have stated, "Another problem is to find convenient and accurate approximations to the distribution of r', for moderate values of n, with close limits of error. A table calculated along the lines suggested in §5 would be very useful." This statement, along with the interest manifested by others in private communications, has led to the investigation reported in this paper.

---

[1] Presented to the American Mathematical Society, December 29, 1936.

[2] Harold Hotelling and Margaret Pabst, *Rank Correlation and Tests of Significance Involving No Assumption of Normality*, Annals of Mathematical Statistics, Vol. VII, 1936, pp. 29-43.

[3] Loc. cit.

II. EXACT DISTRIBUTION OF SUMS OF SQUARED DIFFERENCES

In the paper mentioned above, the authors have given the exact probabilities for all possible values of r' for $n = 2, 3$, and 4. Since r' is a linear function of Σd^2 for any particular value of n, there is a one-to-one correspondence between values of Σd^2 and values of r'. For example, for the case of $n = 3$, we have the following:

$$
\begin{array}{rcccc}
\Sigma d^2 = & 0 & 2 & 6 & 8 \\[6pt]
r' = & 1 & \frac{1}{2} & -\frac{1}{2} & -1 \\[6pt]
p = & \frac{1}{3!} & \frac{2}{3!} & \frac{2}{3!} & \frac{1}{3!}
\end{array}
$$

where p represents the relative frequency of r' or of Σd^2. Therefore it seems pertinent to investigate the distribution of Σd^2 for various values of n.

If n individuals are ranked $1, 2, 3, \cdots n$, by one criterion and then are re-ranked at random there are $n!$ possibilities for the new ranking. Let us consider the differences between the numbers in the new and in the original rankings. Suppose these differences are represented by $d_1, d_2, \cdots d_n$. Then it is apparent that $\sum_{i=1}^{n} d_i^2 = 0$. If we let $a_1, a_2, \cdots a_k$ represent an arrangement for $n = k$, insert $k + 1$ after a_k and advance the cycle one position at a time, we have the following arrangements for the case, $n = k + 1$:

$$
\begin{array}{ccccccccc}
a_1, & a_2, & a_3, & \cdot & \cdot & \cdot & a_k, & k+1 \\
a_2, & a_3, & a_4, & \cdot & \cdot & \cdot & k+1, & a_1 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
a_k, & k+1, & a_1, & \cdot & \cdot & \cdot & a_{k-2}, & a_{k-1} \\
k+1, & a_1, & a_2, & \cdot & \cdot & \cdot & a_{k-1}, & a_k
\end{array}
\qquad (1)
$$

Now, for $n = k$, $d_1 = a_1 - 1$, $d_2 = a_2 - 2, \cdots d_k = a_k - k$. If we list the differences for the $k + 1$ derived arrangements, we have

$$
\begin{array}{cccccccc}
d_1, & d_2, & d_3, & \cdot & \cdot & \cdot & d_k, & 0 \\
d_2+1, & d_3+1, & d_4+1, & \cdot & \cdot & \cdot & 1, & d_1-k \\
d_3+2, & d_4+2, & d_5+2, & \cdot & \cdot & 2, & d_1+1-k, & d_2+1-k \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
d_k+k-1, & k-1, & d_1-2, & \cdot & \cdot & \cdot & \cdot & d_{k-1}-2 \\
k, & d_1-1, & d_2-1, & \cdot & \cdot & \cdot & \cdot & d_k-1
\end{array}
\qquad (2)
$$

It is apparent that each row of differences is formed as follows: the entry in the first column is formed by adding 1 to the entry in column two in the row above, the entry in the second column is obtained by adding 1 to the entry in the third column in the row above, and so on until we come to the entry in the last column which is obtained by subtracting k from the entry in the first column of the preceding row.

If we form the sum of squares of the entries in each row we observe an interesting property of the set; the sums are all congruent, modulus $(k + 1)$. Let us write the sums, denoting them by $S_1, S_2, \cdots S_{k+1}$. Also let d_{ij} represent the entry in the ith row and jth column. Then

$$\begin{aligned} S_i &= \sum_{j=1}^{k+1} d_{i,j}^2 \\ S_{i+1} &= \sum_{j=1}^{k+1} d_{i+1,j}^2 = \sum_{j=2}^{k+1} (d_{i,j} + 1)^2 + (d_{i,1} - k)^2 \\ &= \sum_{j=1}^{k+1} (d_{i,j} + 1)^2 + (d_{i,1} - k)^2 - (d_{i,1} + 1)^2 \\ &= \sum_{j=1}^{k+1} (d_{i,j}^2 + 2d_{i,j} + 1) - (2d_{i,1} - k + 1)(k + 1) \\ &= S_i + 0 + (k + 1) - (2d_{i,1} - k + 1)(k + 1) \\ &= S_i + (k - 2d_{i,1})(k + 1) \end{aligned} \tag{3}$$

Noticing that $d_{i,1} = d_1 + i - 1$, for $i = 1, 2, \cdots k$, and $d_{k+1,1} = k$, we have

$$\begin{aligned} S_2 &= S_1 + (k - 2d_1)(k + 1) \\ S_3 &= S_2 + (k - 2d_2 - 2)(k + 1) \\ S_4 &= S_3 + (k - 2d_3 - 4)(k + 1) \\ &\;\;\vdots \\ S_{k+1} &= S_k + (k - 2d_k - 2k + 2)(k + 1) \\ S_{k+2} &= S_{k+1} + (k - 2k)(k + 1) = S_{k+1} - k(k + 1) \end{aligned} \tag{4}$$

Of course, $S_{k+2} = S_1$, as the $(k + 2)$nd row is identical with the first and the set is closed. So we may write

$$S_{k+1} = S_1 + k(k + 1) \tag{5}$$

The analysis given above not only establishes the congruence of the sums, modulus $(k + 1)$, but also indicates a method of deriving the sums for $n = k + 1$ from the sums for $n = k$, since $S_1 = \sum_{i=1}^{k} d_i^2$. It is also worth noticing that S_{i+1} depends not only on S_i (and therefore on S_1) but also on $d_{i,1}$ (and therefore on d_{i-1}).

Another matter needs attention. It is the relation between the sums of squares of deviations for a particular order and for the reverse order. Let $a_1, a_2, \cdots a_n$ be a particular arrangement. Then the reverse order is a_n, $a_{n-1}, \cdots a_1$. The sums of the squares of the deviates are, respectively,

$$\begin{aligned} S &= (a_1 - 1)^2 + (a_2 \quad - 2)^2 + \cdots (a_k - k)^2 \\ \text{and} \qquad \bar{S} &= (a_k - 1)^2 + (a_{k-1} - 2)^2 + \cdots (a_1 - k)^2 \end{aligned} \tag{6}$$

Then

$$S + \bar{S} = [(a_1 - 1)^2 + (a_1 - k)^2] + [(a_2 - 2)^2 + (a_2 - k + 1)^2] + \cdots [(a_k - k)^2 + (a_k - 1)^2]$$

$$= \sum_{r=1}^{k} [(a_r - r)^2 + (a_r - k + r - 1)^2]$$

$$= \sum_{r=1}^{k} [(a_r - r) + (a_r - k + r - 1)]^2 - 2 \sum_{r=1}^{k} (a_r - r)(a_r - k + r - 1)$$

$$= 4 \sum_{r=1}^{k} a_r^2 - 4(k+1) \sum_{r=1}^{k} a_r + (k+1) \sum_{r=1}^{k} 1 - 2 \sum_{r=1}^{k} a_r^2 + 2(k+1) \sum_{r=1}^{k} a_r - 2(k+1) \sum_{r=1}^{k} r + 2 \sum_{r=1}^{k} r^2.$$

Noting that $\Sigma a_r^2 = \Sigma r^2$ and $\Sigma a_r = \Sigma r$, we readily obtain the result[4]

$$S + \bar{S} = \frac{k^3 - k}{3} \tag{7}$$

It is now apparent that the sums range from 0 to $\frac{k^3 - k}{3}$ with a mean of $\frac{k^3 - k}{6}$.

As the exact frequencies for sums of squares do not seem to be available, it seems useful to compute them for certain small values of n and, at the same time

[4] The geometric representation of the problem may be of some interest. Let the coordinates of point R, in Euclidean n-space be $(1, 2, 3, \cdots n)$, the coordinates of $\bar{R}$ be $(n, n-1, \cdots 2, 1)$, and the coordinates of P be $(x_1, x_2, \cdots x_n)$. Let us restrict the x's to be the numbers $(1, 2, 3, \cdots n)$, but not necessarily in the order given, i.e. the locus of P is a set of $n!$ points, corresponding to the permutations of the numbers $1, 2, 3, \cdots n$. Then it is easy to see that $\sum_{i=1}^{n} x_i = \frac{n^2 + n}{2}$ and that points P lie on an n-flat or hyperplane. Also $\sum_{i=1}^{n} x_i^2 = \frac{n(n+1)(2n+1)}{6}$ so points P lie on a hypersphere with center at the origin. Let us consider the joins PR and $P\bar{R}$. It is readily established that they are orthogonal. Then $(PR)^2 + (P\bar{R})^2 = (R\bar{R})^2 = \frac{n^3 - n}{3}$ or, since $S = \sum_{i=1}^{n} (x_i - i)^2$ and $\bar{S} = \sum_{i=1}^{n} (x_i - n + i - 1)^2$, $S + \bar{S} = \frac{n^3 - n}{3}$ a result previously established otherwise.

to devise a method which can be used successfully to extend the computation to larger values of n if desired. The details of the method follow.

Let D_n represent any series of n differences, $d_1, d_2, \cdots d_n$, and let O_n be an operator such that O_n operating on D_n (written $O_n(D_n)$) means that $D_n = (d_1, d_2, \cdots d_n)$ is changed to $(d_2 + 1, d_3 + 1 \cdots d_n + 1, d_1 - (n - 1))$. Let $\underline{m}$, written following $d_1, d_2, \cdots d_n$, indicate that $\Sigma d^2 = m$. For $n = 3$

$$D_{3,1} = (0, \quad 0, \quad 0):\underline{0}$$

$$O_3(D_{3,1}) = D_{3,2} = (1, \quad 1, -2):\underline{6}$$

$$O_3(D_{3,2}) = D_{3,3} = (2, -1, -1):\underline{6}$$

But we have shown that $S + \bar{S} = \dfrac{k^3 - k}{3}$ for $n = k$. Therefore, for $n = 3$, we have $S + \bar{S} = 8$, so sums of 0 and 6 indicate corresponding sums of 8 and 2 when the order of the elements is reversed. Thus we have, for $n = 3$.

Sums	0	2	4	6	8
Frequencies	1	2	0	2	1

For $n = 4$ we have

$$D_{4,1,1} = (0, \quad 0, \quad 0, \quad 0)$$

$$D_{4,2,1} = (1, \quad 1, -2, \quad 0)$$

$$D_{4,3,1} = (2, -1, -1, \quad 0)$$

where these are obtained from $D_{3,1}$, $D_{3,2}$ and $D_{3,3}$ respectively by inserting a zero as a fourth difference. We operate on each of these four times with O_4. For example,

$$D_{4,2,1} = (1, \quad 1, -2, \quad 0): \underline{6}$$

$$O_4(D_{4,2,1}) = D_{4,2,2} = (2, -1, \quad 1, -2):\underline{10}$$

$$O_4(D_{4,2,2}) = D_{4,2,3} = (0, \quad 2, -1, -1): \underline{6}$$

$$O_4(D_{4,2,3}) = D_{4,2,4} = (3, \quad 0, \quad 0, -3):\underline{18}$$

$$O_4(D_{4,2,4}) = D_{4,2,1} = (1, \quad 1, -2, \quad 0)$$

As a check on computation, we notice, first, that the set is closed by the reappearance of $D_{4,2,1}$; and, second, that 10, 6, 18 and 6 are congruent, modulus 4. In like fashion, one of the sets for $n = 5$, is the following;

$$D_{5,2,4,1} = (3, \quad 0, \quad 0, -3, \quad 0):\underline{18}$$

$$D_{5,2,4,2} = (1, \quad 1, -2, \quad 1, -1): \underline{8}$$

$$D_{5,2,4,3} = (2, -1, \quad 2, \quad 0, -3):\underline{18}$$

$$D_{5,2,4,4} = (0, \quad 3, \quad 1, -2, -2):\underline{18}$$

$$D_{5,2,4,5} = (4, \quad 2, -1, -1, -4):\underline{38}$$

$$D_{5,2,4,1} = (3, \quad 0, \quad 0, -3, \quad 0)$$

Of course the sums for $n = 5$ can be obtained from those for $n = 4$ by making use of (4). For $D_{4,2,4} = (3, 0, 0, -3):\underline{18}$ we have $S_1 = 18$, $k = 4$, $d_1 = 3$, $d_2 = 0$, $d_3 = 0$, $d_4 = -3$. Then

$$S_1 = 18$$

$$S_2 = S_1 + (4 - 2\cdot 3)(5) = 8$$

$$S_3 = S_2 + (4 - 2\cdot 0 - 2)(5) = 18$$

$$S_4 = S_3 + (4 - 2\cdot 0 - 4)(5) = 18$$

$$S_5 = S_4 + (4 - 2\cdot -3 - 6)(5) = 38$$

$$S_1 = S_5 - 4\cdot 5 = 18.$$

However, results obtained by this latter method do not help with the case of $n = 6$. If we desire to obtain results for $n = 6$ we will need to exhibit the complete sets of differences for $n = 5$ as we did by the former method.

An alternative method for obtaining frequencies of sums of squares is of some interest. It will be illustrated for $n = 4$. Let us consider the square array

$$\begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{pmatrix}$$

If we form all possible products $a_i b_j c_k d_l (i, j, k, l = 1, 2, 3, 4; i \neq j \neq k \neq l)$, the subscripts give the $4!$ permutations of 1, 2, 3, 4. Now let us form a new array

$$\begin{pmatrix} a_0 & b_1 & c_2 & d_3 \\ a_{-1} & b_0 & c_1 & d_2 \\ a_{-2} & b_{-1} & c_0 & d_1 \\ a_{-3} & b_{-2} & c_{-1} & d_0 \end{pmatrix}$$

where subscripts in each column represent the vertical distance of the term above the principal diagonal. Since the original terms had subscripts giving all possible arrangements of 1, 2, 3, 4, terms formed in a similar fashion from the new array will give all possible arrangements of the differences. Now form a third array

$$\begin{pmatrix} x^0 & x^1 & x^4 & x^9 \\ x^1 & x^0 & x^1 & x^4 \\ x^4 & x^1 & x^0 & x^1 \\ x^9 & x^4 & x^1 & x^0 \end{pmatrix}$$

where the exponent of x is the square of the corresponding subscript in the

TABLE I

Frequencies of sums of squares of rank differences

Σd^2 \ N	2	3	4	5	6	7
0	1	1	1	1	1	1
	*					
2	1	2	3	4	5	6
4		*0	1	3	6	10
6		2	4	6	9	14
8		1	2	7	16	29
10			*2	6	12	26
12			2	4	14	35
14			4	10	24	46
16			1	6	20	55
18			3	10	21	54
20			1	*6	23	74
22				10	28	70
24				6	24	84
26				10	34	90
28				4	20	78
30				6	32	90
32				7	42	129
34				6	29	106
					*	
36				3	29	123
38				4	42	134
40				1	32	147
42					20	98
44					34	168
46					24	130
48					28	175
50					23	144
52					21	168
54					20	144
56					24	*184

* The asterisk shows the location of the mean. The frequencies for $n = 6, 7$ extend beyond the limits of the table but may easily be obtained by symmetry.

second array. It is easy to see that, if terms are formed from the new array by the same method as before, our terms are powers of x where the exponents represent sums of squares of differences. If we now define the array to be equal to the sum of the terms formed from the array, then

$$\begin{pmatrix} x^0 & x^1 & x^4 & x^9 \\ x^1 & x^0 & x^1 & x^4 \\ x^4 & x^1 & x^0 & x^1 \\ x^9 & x^4 & x^1 & x^0 \end{pmatrix} = k_1x^0 + k_2x^2 + \cdots k_6x^{10} + \cdots k_2x^{18} + k_1x^{20},$$

and the k's give the desired frequencies for sums of squares corresponding to exponents of x. For example $\Sigma d^2 = 0$ occurs k_1 times, $\Sigma d^2 = 2$ occurs k_2 times, etc.

It can be readily verified that, for $n < 5$, the array can be expanded as a determinant and the values of the k's can be obtained by taking the absolute values of the coefficients in the expansion. Also, considering the arrays as determinants, their values for $n = 2, 3, 4$ are, respectively, $(1 - x^2)$, $(1 - x^2)^2 (1 - x^4)$, $(1 - x^2)^3 (1 - x^4)^2 (1 - x^6)$. If it were possible to obtain a general form of this type it might be possible to greatly reduce the labor which is involved in expanding the arrays. At present, however, this method of attack does not seem feasible on account of the lack of adequate sub-checks, the amount of work involved, and its inappropriateness for use by inexperienced clerical help.

Hotelling and Pabst[5] have given exact results in terms of n for the cases $\Sigma d^2 = 0, 2, 4, 6$. It is certainly possible to follow their method to obtain general results for Σd^2 larger than 6, but, as they suggest, the work becomes very laborious. For $\Sigma d^2 = 8$ we need the sets of possible integral values for $x_1, x_2, \cdots x_n$, under the following conditions: (a) $\sum_{i=1}^{n} x_i = 0$, (b) $\sum_{i=1}^{n} x_i^2 = 8$, (c) $1 + x_1, 2 + x_2, 3 + x_3, 4 + x_4, \cdots n + x_n$ are the numbers $1, 2, 3, \cdots n$, (but not necessarily in that order).

Possible solutions are:

(a) $x_{i-2} = 2, x_{i-1} = 0, x_i = -2$ ($i = 3, 4, \cdots n$) and the other x's zero,

(b) $x_{b-2} = 2, x_{b-1} = -1, x_b = -1, x_{a-1} = 1, x_a = -1$ ($a = 5, 6, \cdots n$; $b = 3, 4, \cdots a - 2$),

(c) $x_{b-1} = 1, x_b = -1, x_{a-2} = 2, x_{a-1} = -1, x_a = -1$ ($a = 5, 6, \cdots n$; $b = 2, 3, \cdots a - 3$),

(d) $x_{b-2} = -2, x_{b-1} = 1, x_b = 1, x_{a-1} = 1, x_a = -1$ ($a = 5, 6, \cdots n$; $b = 3, 4, \cdots a - 2$),

(e) $x_{b-1} = 1, x_b = -1, x_{a-2} = -2, x_{a-1} = 1, x_a = 1$ ($a = 5, 6, \cdots n$; $b = 2, 3, \cdots a - 3$),

(f) $x_{d-1} = x_{c-1} = x_{b-1} = x_{a-1} = 1; x_d = x_c = x_b = x_a = -1$ ($a = 8, 9, \cdots n$; $b = 6, 7, \cdots a - 2; c = 4, 5, \cdots b - 2; d = 2, 3, \cdots c - 2$)

Frequencies for each of these types must be considered separately. The

[5] Loc. cit. p. 35.

method of evaluation will be illustrated for type (f), since this type yields the polynomial of highest degree. It is apparent that the required frequency is obtained by computing $\sum_{8}^{n}\left(\sum_{6}^{a-2}\left(\sum_{4}^{b-2}\left(\sum_{2}^{c-2}1\right)\right)\right)$. It can be verified that the result is

$$\frac{(n-4)^{(4)}}{4!} = \frac{(n-4)(n-5)(n-6)(n-7)}{24}$$

The total of (a), (b), (c), (d), (e), and (f) is

$$(n-2) + 2(n-3)^{(2)} + \frac{(n-4)^{(4)}}{4!}$$

For $\Sigma d^2 = 10$, the result seems to be

$$2(n-3) + (n-3)^{(2)} + (n-4)^{(3)} + \frac{(n-5)^{(5)}}{5!}.$$

For sums greater than 8 the method becomes quite uninviting, not only because of the intricacy of the necessary analysis, but also because of the opportunities for mechanical errors and the absence of satisfactory checks. Besides, if the exact distribution for a particular value of n is desired, we need expressions for $\Sigma d^2 = 0, 2, 4, \cdots \frac{n^3-n}{6} - 2$. For n as small as 8, this means the requirement of 42 formulas. It is fairly evident that these formulas will comprise polynomials ranging in degree from 0 to 41.

III. APPROXIMATIONS

Since the exact distributions of sums of squares are not easily obtained, we next consider the problem of finding approximations for them. Hotelling and Pabst[6] have given a method of deriving the even moments of the distribution of r', (the odd moments being zero), and have recorded the values of the second and fourth moments. They have also remarked that the kurtosis, $\beta_2 = \mu_4/\mu_2^2$, approaches 3 and that the distribution of r' approaches normality as n approaches infinity. These are valuable and interesting results. Because of them the normal curve suggests itself as an approximating function. Its use has been considered a little later in this investigation.

But a distribution with a finite range causes trouble at the tails when a normal fit is attempted, and, for this problem, we are particularly interested in the tails. It seems more feasible to attempt an approximation with the Pearson type II curve, $y = y_0\left(1 - \frac{x^2}{a^2}\right)^m$. This has the advantage of a finite range and three

[6] Loc. cit. p. 32 et seq.

constants to be determined. The values of these constants, as given by Elderton[7] are

$$m = \frac{5\beta_2 - 9}{2(3 - \beta_2)}, \qquad a^2 = \frac{2\mu_2\beta_2}{3 - \beta_2}, \qquad y_0 = \frac{N \times \Gamma\,(2m + 2)}{a \times 2^{2m+1} \times [\Gamma\,(m + 1)]^2} \quad (8)$$

(where N is the total frequency).

If we use this distribution to approximate the distribution of sums of squares, it proves convenient to define x as equal to one-half the deviation of Σd^2 from its mean, i.e.,

$$x = \frac{\Sigma d^2}{2} - \frac{n^3 - n}{12}$$

Then the relative frequency of $\Sigma d^2 = k$ is approximated by

$$\int_{x_s - \frac{1}{2}}^{x_s + \frac{1}{2}} f(x)\,dx \doteq f(x_s) \qquad \text{where } x_s = \frac{k}{2} - \frac{n^3 - n}{12}$$

(Of course, closer approximations may be obtained, if desired). The approximation used is clear if we remember that only even values of k are possible and that the range is now $\frac{n^3 - n}{6}$.

The moments for x are now obtained from the moments for r' by multiplying by the proper powers of $\frac{n^3 - n}{12}$. We have

$$\mu_2(x) = (n - 1)\left[\frac{n(n + 1)}{6}\right]^2$$

The value of β_2 is unchanged. For r' or x it is

$$\beta_2 = \frac{3(25n^4 - 13n^3 - 73n^2 + 37n + 72)}{25n(n + 1)^2(n - 1)}$$

For $n = 5$, $\mu_2 = 25$, $\beta_2 \doteq 2.0720$, $N = 51$. Using these values and equations (8), we obtain $a = 10.566$, $m = .73276$, $y_0 = 7.8545$. The approximating function is $y = 7.8545\left(1 - \frac{x^2}{111.64}\right)^{.73276}$ In table II the computed values of y and the true frequencies are listed for comparison.

When testing the significance of a particular value of Σd^2 our principal interest is in the probability that $\Sigma d^2 \leq k$, rather than in the probability that $\Sigma d^2 = k$. The probability that $\Sigma d^2 \leq k$ requires cumulation of frequencies, followed by division by the total frequency. If results, given in table II, are compared it is noticed that the maximum error in using the type II function is .0194 and the average error is .0072. Comparisons for other values of n are given in table III.

[7] Elderton, W. P., *Frequency Curves and Correlation*, Layton, London, 2nd ed., 1927, p. 84.

TABLE II

Comparison of exact and approximate frequencies for $n = 5$

$$\left(\text{Approximations obtained by computing ordinates of } y = 7.845\left(1 - \frac{x^2}{111.64}\right)^{.73276}\right)$$

Σd^2	Frequencies		Cumulative (expressed as percent of 120)		Difference of cumulatives
	Exact	Approximate	Exact	Approximate	
0	1	1.50	.0083	.0125	−.0042
2	4	3.04	.0417	.0378	+.0039
4	3	4.21	.0667	.0729	−.0062
6	6	5.14	.1167	.1157	+.0010
8	7	5.91	.1750	.1650	+.0100
10	6	6.52	.2250	.2193	+.0057
12	4	7.01	.2583	.2777	−.0194
14	10	7.39	.3417	.3393	+.0024
16	6	7.65	.3917	.4031	−.0114
18	10	7.80	.4750	.4681	+.0069
20	6	7.85	.5250	.5335	−.0085
					average of absolute values = .0072

TABLE III

Approximating functions, with errors involved

n	Approximating functions		Average and maximum absolute values of differences of cumulatives					
	Type II	Normal	Exact—type II		Exact—normal		Type II—normal	
5	$7.8545\left(1 - \frac{x^2}{111.64}\right)^{.73276}$	$\frac{5!}{\sqrt{50\pi}} e^{-\frac{x^2}{50}}$	.0072	.0194	.0200	.0415	.0210	.0357
6	$31.652\left(1 - \frac{x^2}{351.75}\right)^{1.3715}$	$\frac{6!}{\sqrt{122.5\pi}} e^{-\frac{x^2}{122.5}}$	.0030	.0126	.0131	.0273	.0136	.0270
7	$156.33\left(1 - \frac{x^2}{918.84}\right)^{2.0160}$	$\frac{7!}{\sqrt{261.33\pi}} e^{-\frac{x^2}{261.33}}$	.0017	.0067	.0106	.0221	.0108	.0209
8	$918.72\left(1 - \frac{x^2}{2098.4}\right)^{2.6635}$	$\frac{8!}{\sqrt{504\pi}} e^{-\frac{x^2}{504}}$					.0086	.0175
9	$6276.3\left(1 - \frac{x^2}{4332.6}\right)^{3.3140}$	$\frac{9!}{\sqrt{900\pi}} e^{-\frac{x^2}{900}}$						
10	$64515\left(1 - \frac{x^2}{8266.6}\right)^{3.9655}$	$\frac{10!}{\sqrt{1512.5\pi}} e^{-\frac{x^2}{1512.5}}$						

It would be very convenient if the cumulative frequencies could be approximated by the use of normal curves. In table III are listed the proper normal curves, along with comparisons with the exact values and with the values obtained from the type II curves. For the values of n investigated the normal curve is not as satisfactory as the type II. This, of course, is to be expected because of the lack of agreement between the fourth moment of the normal curve and of the exact distribution. However, in view of the fact that, for values of n investigated, the maximum and average errors decrease as n increases, it seems satisfactory to sacrifice accuracy to expedience and use the normal curve as an approximating function for cases of n greater than 10. This has been done in constructing table V. In further justification it might be noted that β_2, which approaches 3 as n approaches infinity is an increasing function of n for n greater than 3.

IV. TABLES TO TEST THE SIGNIFICANCE OF THE RANK CORRELATION COEFFICIENT, WITH EXAMPLES OF THEIR USE

Table IV gives the probability that, for any given value of n and a computed value of Σd^2 less than or equal to the mean, the value will not be exceeded by chance. For a value of Σd^2 greater than or equal to the mean, it gives the probability that the value will be equalled or exceeded. The values for $n = 2, 3, 4, 5, 6, 7$ are computed from exact frequencies; those for $n = 8, 9, 10$ are computed from type II curves.

Table V is constructed by the use of normal curves. It gives the limits of Σd^2 for a few of the more useful probabilities.

It seems desirable to explain why values of Σd^2 were tabled rather than values of r'. It was done for two reasons: first, to avoid the difficulties arising from discrete variates; and, second, because the tables seem more useful in the form given since the labor of completing the calculation of r' can be avoided if the computed value of Σd^2 tests as not significant.

Example 1. Seven individuals are ranked by two criteria, as indicated below. Are the results significantly alike?

A	1	2	3	4	5	6	7	
B	2	1	6	3	4	7	5	
d	1	−1	3	−1	−1	1	−2	/ 0
d^2	1	1	9	1	1	1	4	/ 18

Solution: Rows 3 and 4 give the differences and squared differences, respectively. If we enter table IV with $n = 7$ and $\Sigma d^2 = 18$, we find $P = .0548$, so we would expect that a value as small as 18 would occur by chance more than 5% of the time. This does not usually indicate significance so it is useless to compute the value of r'. It is interesting to notice that r' actually does prove to be equal to .68 and that, if we had used the formula, $\sigma_{r'} = 1.0471 \left(\frac{1 - r'^2}{\sqrt{n}}\right)$ we might have

TABLE IV

The probability that $\Sigma d^2 \geq S$ for $S \geq \Sigma_M$, or that $\Sigma d^2 \leq S$ for $S \leq \Sigma_M$ (where Σ_M represents mean value of sum of squares)

	$N = 2$	3	4	5	6	7	8	9	10
Σ_M	1	4	10	20	35	56	84	120	165
S									
0	.5000	.1667	.0417	.0083	.0014	.0002	.0000	.0001	.0000
2	.5000	.5000	.1667	.0417	.0083	.0014	.0006	.0002	.0001
4		.5000	.2083	.0667	.0167	.0034	.0011	.0003	.0001
6		.5000	.3750	.1167	.0292	.0062	.0018	.0005	.0001
8		.1667	.4583	.1750	.0514	.0119	.0028	.0007	.0002
10			.5417	.2250	.0681	.0171	.0042	.0010	.0003
12			.4583	.2583	.0875	.0240	.0059	.0015	.0004
14			.3750	.3417	.1208	.0331	.0081	.0020	.0005
16			.2083	.3917	.1486	.0440	.0108	.0027	.0007
18			.1667	.4750	.1778	.0548	.0141	.0035	.0009
20			.0417	.5250	.2097	.0694	.0179	.0045	.0011
22				.4750	.2486	.0833	.0224	.0057	.0014
24				.3917	.2819	.1000	.0275	.0071	.0018
26				.3417	.3292	.1179	.0331	.0087	.0022
28				.2583	.3569	.1333	.0396	.0106	.0027
30				.2250	.4014	.1512	.0469	.0127	.0032
32				.1750	.4597	.1768	.0550	.0152	.0039
34				.1167	.5000	.1978	.0639	.0179	.0046
36				.0667	.5000	.2222	.0736	.0210	.0054
38				.0417	.4597	.2488	.0841	.0244	.0064
40				.0083	.4014	.2780	.0956	.0281	.0075
42					.3569	.2974	.1078	.0323	.0086
44					.3292	.3308	.1207	.0368	.0100
46					.2819	.3565	.1345	.0417	.0114
48					.2486	.3913	.1491	.0470	.0130
50					.2097	.4198	.1645	.0528	.0148
52					.1778	.4532	.1806	.0589	.0168
54					.1486	.4817	.1974	.0656	.0189
56					.1208	.5183	.2150	.0726	.0212
58					.0875	.4817	.2332	.0802	.0237

TABLE IV—*Continued*

N	6	7	8	9	10
Σ_M	35	56	84	120	165
S					
60	.0681	.4532	.2520	.0882	.0264
62	.0514	.4198	.2715	.0966	.0293
64	.0292	.3913	.2915	.1056	.0324
66	.0167	.3565	.3120	.1149	.0358
68	.0083	.3308	.3330	.1248	.0394
70	.0014	.2974	.3544	.1351	.0432
72		.2780	.3761	.1459	.0472
74		.2488	.3982	.1571	.0515
76		.2222	.4205	.1688	.0561
78		.1978	.4431	.1809	.0609
80		.1768	.4657	.1936	.0659
82		.1512	.4885	.2065	.0713
84		.1333	.5113	.2198	.0769
86		.1179	.4885	.2336	.0828
88		.1000	.4657	.2477	.0889
90		.0833	.4431	.2622	.0954
92		.0694	.4205	.2770	.1021
94		.0548	.3982	.2922	.1091
96		.0440	.3761	.3077	.1164
98		.0331	.3544	.3234	.1239
100		.0240	.3330	.3394	.1318
102		.0171	.3120	.3557	.1399
104		.0119	.2915	.3721	.1483
106		.0062	.2715	.3888	.1570
108		.0034	.2520	.4056	.1659
110		.0014	.2332	.4226	.1751
112		.0002	.2150	.4397	.1846
114			.1974	.4568	.1944
116			.1806	.4741	.2044
118			.1645	.4914	.2146
120			.1491	.5086	.2251
122			.1345	.4914	.2358
124			.1207	.4741	.2468

TABLE IV—*Concluded*

N	8	9	10	N	8	9	10
Σ_M	84	120	165	Σ_M	84	120	165
S				S			
126	.1078	.4568	.2580	168	.0003	.1459	.4865
128	.0956	.4397	.2694	170		.1351	.4731
130	.0841	.4226	.2810	172		.1248	.4596
132	.0736	.4056	.2928	174		.1149	.4462
134	.0639	.3888	.3048	176		.1056	.4328
136	.0550	.3721	.3169	178		.0966	.4196
138	.0469	.3557	.3293	180		.0882	.4063
140	.0396	.3394	.3418	182		.0802	.3931
142	.0331	.3234	.3545	184		.0726	.3802
144	.0275	.3077	.3673	186		.0656	.3673
146	.0224	.2922	.3802	188		.0589	.3545
148	.0179	.2770	.3932	190		.0528	.3418
150	.0141	.2622	.4063	200		.0470	.3293
152	.0108	.2477	.4196	202		.0417	.3169
154	.0081	.2336	.4328	204		.0368	.3048
156	.0059	.2198	.4462	206		.0323	.2928
158	.0042	.2065	.4596	208		.0281	.2810
160	.0028	.1935	.4731	210		.0244	.2694
162	.0018	.1809	.4865	212		.0210	.2580
164	.0011	.1688	.5000	214		.0179	.2468
166	.0006	.1571	.5000	216		.0152	.2358

(Tables for cases 9 and 10 can be completed by symmetry.)

judged the value of r' significant, since $\sigma_{r'} = .213$, and .213 is less than one-third of .68.

Example 2. Six golfers found, upon ranking their scores and also ranking their respective amounts of sleep for the previous night, that the two orders were the reverse of one another except that the two ranking 1, 2 in sleep ranked 5, 6 in score. Is the negative correlation too great to be reasonably attributed to chance?

Solution: We find $\Sigma d^2 = 68$ and, upon consulting table IV, $P = .0083$, so we conclude that more sleep might mean fewer strokes.

Example 3. Before an examination a teacher ranked his class of 13 members.

After the examination he found that the sum of the squares of the deviations of rank on examination from rank estimated was 144. Should he consider the agreement satisfactory?

TABLE V

Pairs of values between which Σd^2 has a probability, P, of being included

N	P = .99		.98		.96		.90		.80	
11	40.8	399.2	58.2	381.8	77.1	362.9	105.6	334.4	130.8	309.2
12	60.9	505.1	82.4	483.6	105.9	460.1	141.2	424.8	172.5	393.5
13	93.3	634.7	119.6	608.4	148.2	579.8	191.2	536.8	229.3	498.7
14	125.9	780.1	161.4	748.6	195.8	714.2	247.4	662.6	293.3	616.7
15	174.5	945.5	211.8	908.2	252.6	867.4	313.8	806.2	368.2	751.8
16	227.8	1132.2	271.6	1088.4	319.4	1040.6	391.2	968.8	455.0	905.0
17	290.5	1341.4	341.4	1290.6	397.0	1235.0	480.4	1151.6	554.6	1077.4
18	363.6	1574.4	422.3	1515.7	486.3	1451.7	582.4	1355.6	667.8	1270.2
19	447.9	1832.1	514.9	1765.1	588.2	1691.8	698.0	1582.0	795.6	1484.4
20	544.1	2115.9	620.2	2039.8	703.4	1956.6	828.1	1831.9	939.0	1721.0
21	653.0	2427.0	738.9	2341.1	832.8	2247.2	973.6	2106.4	1098.7	1981.3
22	775.5	2766.5	872.0	2670.0	977.3	2564.7	1135.3	2406.7	1275.7	2266.3
23	912.5	3135.5	1020.2	3027.8	1137.8	2910.2	1314.2	2733.8	1471.0	2577.0
24	1064.7	3535.3	1184.3	3415.7	1315.1	3284.9	1511.1	3088.9	1685.4	2914.6
25	1233.0	3967.0	1365.4	3834.6	1510.1	3689.9	1727.0	3473.0	1919.8	3280.2
26	1418.2	4431.8	1564.1	4285.9	1723.6	4126.4	1962.7	3887.3	2175.3	3674.7
27	1621.1	4930.9	1781.5	4770.5	1956.5	4595.5	2219.2	4332.8	2452.6	4099.4
28	1842.7	5465.3	2018.1	5289.9	2209.8	5098.2	2497.3	4810.7	2752.8	4555.2
29	2083.7	6036.3	2275.1	5744.9	2484.3	5635.7	2797.9	5322.1	3076.7	5043.3
30	2345.0	6645.0	2553.2	6436.8	2780.8	6209.2	3122.0	5868.0	3425.2	5564.8

Solution: Entering table V with $n = 13$ we see that $P = .96$ for a value between 148.2 and 579.8, and that $P = .98$ for a value between 119.6 and 608.4. Therefore the probability of not exceeding 144 by chance is between .02 and .01. It would seem that the teacher showed considerable knowledge of his class.

CARNEGIE INSTITUTE OF TECHNOLOGY.

NOTE ON CORRELATIONS

By D. B. De Lury

When the value of a correlation coefficient is to be estimated from a set of N pairs of observations, (x_i, y_i), $i = 1, 2, \cdots N$, the statistic ordinarily computed is, of course, the product-moment correlation coefficient,

$$r = s_{12}/(s_1 s_2), \quad \text{where}$$

$$ns_1^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2, \qquad ns_2^2 = \sum_{i=1}^{N} (y_i - \bar{y})^2, \qquad ns_{12} = \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}),$$

$$N\bar{x} = \sum_{i=1}^{N} x_i, \qquad N\bar{y} = \sum_{i=1}^{N} y_i, \qquad n = N - 1.$$

However, when x and y are known to have the same population mean and variance, the precision of the estimate may be improved slightly by using the intraclass correlation coefficient,

$$r' = \frac{2\sum_{1}^{N} (x_i - \xi)(y_i - \xi)}{\sum_{1}^{N} \{(x_i - \xi)^2 + (y_i - \xi)^2\}}, \qquad 2N\xi = \sum_{1}^{N} (x_i + y_i).$$

It may be of interest to inquire into the properties of an analogous coefficient, appropriate to the case of equal variances and different means. This coefficient would naturally be chosen to be

$$u = 2s_{12}/(s_1^2 + s_2^2) = \{2s_1 s_2/(s_1^2 + s_2^2)\}r.$$

Obviously, $|u| \leq |r|$.

The probability distribution of u is easily determined, under the assumption that x and y obey a bivariate normal distribution. If σ^2 is their common variance, no restriction is introduced by taking $\sigma = 1$. Then the probability element of s_1, s_2, r, is known to be[1]

$$\frac{n^n}{\pi(n-2)!(1-\rho^2)^{n/2}} (s_1 s_2)^{n-1} e^{-\frac{n}{2(1-\rho^2)}(s_1^2 - 2\rho r s_1 s_2 + s_2^2)} (1 - r^2)^{\frac{n-3}{2}} ds_1 ds_2 dr,$$

where ρ is the correlation of x and y. From this, the distribution of u can be obtained by making the transformation

$$u = \{2s_1 s_2/(s_1^2 + s_2^2)\}r, \qquad v = 2s_1 s_2/(s_1^2 + s_2^2), \qquad w = s_1^2 + s_2^2.$$

[1] R. A. Fisher, Biometrika, Vol. 10, p. 510.

Under this transformation, the range of s_1, s_2, r, determined by the inequalities $0 \leq s_i \leq \infty$, $i = 1, 2$, $-1 \leq r \leq 1$, is mapped in a two-fold manner upon the space $-v \leq u \leq v$, $0 \leq v \leq 1$, $0 \leq w \leq \infty$. For fixed u, v ranges from u to 1 or from $-u$ to 1, according as u is positive or negative, and w runs from 0 to ∞. The probability element of u, v, w, is found to be

$$\frac{(n/2)^n}{\pi(n-2)!(1-\rho^2)^{n/2}} \frac{v}{\sqrt{1-v^2}} (v^2 - u^2)^{(n-3)/2} w^{n-1} e^{-\frac{n}{2(1-\rho^2)}(1-\rho u)w} \, du \, dv \, dw,$$

and the distribution of u, obtained by integrating with respect to v and w, is

$$K(1-\rho^2)^{n/2}(1-\rho u)^{-n}(1-u^2)^{(n-2)/2} du, \qquad K = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)}.$$

If $\rho = 0$, the distribution of u is identical with that of r, the product-moment correlation coefficient (for $\rho = 0$), in samples of $(N + 1)$ pairs of observations. Therefore, to test the hypothesis of independence, using the coefficient u, the methods and tables appropriate to testing the same hypothesis, using the coefficient r, are available. The precision gained by using u rather than r is equivalent to that supplied by another pair of observations.

In the general case, the transformation introduced by R. A. Fisher,[2]

$$u = \tanh z, \; \rho = \tanh \zeta,$$

leads to the distribution element[3]

$$K \operatorname{sech}^n(z - \zeta) dz.$$

This distribution is invariant in form under varying ζ, and is effectively normal for samples of any size. In all cases, z is an unbiased estimate of ζ.

The variance of z can be obtained by the following device. Denote by $I(2p, n)$ the 2 p-th moment of z about the mean,

$$I(2p, n) = K \int_{-\infty}^{\infty} x^{2p} \operatorname{sech}^n x \, dx.$$

Integration by parts gives the recurrence formula,

$$I(2p, n) = \frac{n^2}{(2p+1)(2p+2)} \{I(2p+2, n) - I(2p+2, n+2)\}, \qquad p \geq 0.$$

---

[2] Metron, Vol. 1, N. 4, p. 7.

[3] The distributions of u and z for $n = 1$ have been given by R. A. Fisher, Metron, Vol. 1, N. 4, p. 8.

From this follows at once the relation

$$I(2p + 2, n + 2) = I(2p + 2, 1) - (2p + 1)(2p + 2)\left\{\frac{I(2p, 1)}{1^2} + \frac{I(2p, 3)}{3^2} + \cdots + \frac{I(2p, n)}{n^2}\right\}, \; n \text{ odd},$$

$$= I(2p + 2, 2) - (2p + 1)(2p + 2)\left\{\frac{I(2p, 2)}{2^2} + \frac{I(2p, 4)}{4^2} + \cdots + \frac{I(2p, n)}{n^2}\right\}, \; n \text{ even}.$$

The values of $I(2p + 2, 1)$ and $I(2p + 2, 2)$ can be found without evaluating the integrals, by letting $n \to \infty$. It can be shown that $I(2p, n) = O(n^{-p})$, and hence $\lim_{n\to\infty} I(2p, n) = 0$ for $p > 0$. We obtain

$$I(2p + 2, 1) = (2p + 1)(2p + 2)\left\{\frac{I(2p, 1)}{1^2} + \frac{I(2p, 3)}{3^2} + \frac{I(2p, 5)}{5^2} + \cdots\right\},$$

$$I(2p + 2, 2) = (2p + 1)(2p + 2)\left\{\frac{I(2p, 2)}{2^2} + \frac{I(2p, 4)}{4^2} + \frac{I(2p, 6)}{6^2} + \cdots\right\}.$$

Hence, for all values of n and p, (replacing $n + 2$ by n),

$$I(2p + 2, n) = (2p + 1)(2p + 2)\left\{\frac{I(2p, n)}{n^2} + \frac{I(2p, n + 2)}{(n + 2)^2} + \frac{I(2p, n + 4)}{(n + 4)^2} + \cdots\right\}.$$

Setting $p = 0$ to get the variance,

$$\mu_2 = I(2, n) = 2\left\{\frac{1}{n^2} + \frac{1}{(n + 2)^2} + \frac{1}{(n + 4)^2} + \cdots\right\}.$$

Therefore, making use of the fact that $\int_m^\infty x^{-2}\,dx < \sum_{i=m}^{\infty} i^{-2} < \int_{m-1}^\infty x^{-2}\,dx$, we find that

$$1/n < \mu_2 < 1/(n - 2),$$

and from the numerical values of μ_2 for small values of n, it appears that the approximation $\mu_2 \sim 1/(n - 1)$ is satisfactory in all cases.

In the same way, it can be shown that

$$3/n^2 < \mu_4 < 3/(n - 2)^2.$$

Thus the method of transforming correlations to test for significance, used by R. A. Fisher in connection with both interclass and intraclass correlations, is available here also, and is, in fact, slightly simpler, owing to the absence of bias.

The coefficient u can, of course, be used in all situations where the intraclass coefficient is appropriate, (when the number of observations in each class is two), and conceivably in a small class of other cases as well. The test of significance is simpler using u instead of r', and the loss of precision is negligible.

UNIVERSITY OF TORONTO.

INTERIOR AND EXTERIOR MEANS OBTAINED BY THE METHOD OF MOMENTS

By Edward L Dodd

1. Introduction—The Substitutive Mean. A very general mean based upon substitution was proposed by O. Chisini.[1] Briefly stated, this mean M of data $x_1, x_2, \cdots, x_n$, is a number which satisfies some equation of the form

$$G(M, M, \cdots, M) = G(x_1, x_2, \cdots, x_n). \tag{1}$$

If, now,[2]

$$M = F(x_1, x_2, \cdots, x_n) \tag{2}$$

is an *explicit* expression of M, then for each value c which each of the arguments x_i can take on,

$$F(c, c, \cdots, c) = c; \tag{3}$$

or at least one value of this F is c.

Suppose now that $F(x_1, x_2, \cdots, x_n)$ is *any* function of $x_1, x_2, \cdots, x_n$, defined for at least one set of equal arguments c, and such that whenever defined for equal arguments c, at least one value of $F(c, c, \cdots, c) = c$. Such a function, I have called a *substitutive mean*. Various extensions[3] are immediate, such as the use of integration in place of summation. Indeed, point set functions or functionals may be used.[4] Here I shall supplement (3) by a fairly common convention. If $F(c, c, \cdots, c)$ is not originally defined, but as $x_i \to c$ simultaneously, limit $F(x_1, x_2, \cdots, x_n) = c$, in this case, $F(c, c, \cdots, c)$ will be assigned its limiting value c,—thus establishing continuity.

2. Location and scale as means. The purpose of this paper is to investigate the nature of the means which arise when the well known *Method of Moments* is used to estimate the values of two important parameters—namely, the location κ and the scale α of a frequency function or distribution. These are taken as associated with the variable x of the distribution thus:

$$t' = (x - \kappa)/\alpha. \tag{4}$$

---

[1] O. Chisini, "Sul Concetto di media," *Periodico di Matematico*, Series 4, Vol. 9, (1929), pp. 106-116.

[2] E. L. Dodd, "Internal and External Means Arising from the Scaling of Frequency Functions," These *Annals*, Vol. 8, (1937), pp. 18-20.

[3] For an extension of Chisini's results, see Bruno de Finetti, "Sul Concetto di media," *Giornale dell' Instituto Italiano degli Attuari*, Vol. 2, (1931), pp. 369-396.

[4] E. L. Dodd, "The Chief Characteristic of Statistical Means," Cowles Commission lecture. *Colorado College Publication*, General Series No. 208, (1936), pp. 89-92.

The nature of the distribution is then "specified" by

$$(5) \qquad y = \alpha^{-1}\Phi(t');$$

where Φ may contain other parameters, but in Φ the κ and α appear only in the t' given by (4). For this mode of approach, the reader is referred to R. A Fisher.[5]

The other parameters which *may* appear in Φ will not be considered in this paper.

The parameters κ and α are in general unknown and unknowable. However, we attempt to get close *estimates* k and a, of κ and α respectively, from a set of observations

$$(6) \qquad X_1, X_2, \cdots, X_n.$$

To accomplish this, we have to solve certain equations formed in some way from

$$(7) \qquad t = (x - \mathrm{k})/\mathrm{a},$$

and

$$(8) \qquad y = \mathrm{a}^{-1}\Phi(t).$$

These equations (7) and (8) result from (4) and (5) by substituting estimates k and a, respectively, for parameters κ and α

Now the Method of Moments equates the theoretic moments—those obtained from some such equation as (8) with t replaced by its value in (7)—to the moments obtained from the observation (6).

For the following discussion it will be useful to obtain "*auxiliary*" moments from the $\Phi(t)$ in (8) *before* substitution is made from (7). Such moments, then, *do not depend* at all upon the values ultimately assigned to k and a. It is supposed that

$$(9) \qquad \int_{-\infty}^{\infty} \Phi(t)\, dt = 1,$$

so that $\Phi(t)$ gives probability or relative frequency Here, for finite distributions, $\Phi(t) = 0$ outside the interval of the distribution We shall assume the existence of the first moment

$$(10) \qquad \mu = \int_{-\infty}^{\infty} t\Phi(t)\, dt,$$

and of the variance

$$(11) \qquad \sigma^2 = \int_{-\infty}^{\infty} (t - \mu)^2\Phi(t)\, dt = \int_{-\infty}^{\infty} t^2\Phi(t)\, dt - \mu^2;$$

and we shall assume that $\sigma > 0$, to eliminate a degenerate case

[5] R. A. Fisher, "On the Mathematical Foundation of Theoretical Statistics" *Philosophical Transactions of the Royal Society of London*, Series A, Vol 222, (1921), pp. 309–368.

For the empirical moments of (6), we write

$$\bar{X} = (X_1 + X_2 + \cdots + X_n)/n = \Sigma X_i/n, \tag{12}$$

$$S = \Sigma X_i^2/n = \Sigma(X_i - \bar{X})^2/n + \bar{X}^2. \tag{13}$$

These two moments are, by the Method of Moments, equated respectively to

$$M_1 = \frac{1}{a}\int_{-\infty}^{\infty} x\Phi[(x - k)/a]\,dx, \tag{14}$$

$$M_2 = \frac{1}{a}\int_{-\infty}^{\infty} x^2\Phi[(x - k)/a]\,dx. \tag{15}$$

But, from (7), (10), and (11) it is easy to see that

$$M_1 = k + a\mu, \tag{16}$$

$$M_2 = k^2 + 2ka\mu + a^2(\sigma^2 + \mu^2); \tag{17}$$

from which it follows that

$$M_2 - M_1^2 = a^2\sigma^2 \tag{18}$$

Suppose now that τ^2 is the empirical variance,

$$\tau^2 = \Sigma(X_i - \bar{X})^2/n. \tag{19}$$

It follows from (12) and (13), that if $M_1 = \bar{X}$ and $M_2 = S$—as the Method of Moments requires—then

$$a^2 = \tau^2/\sigma^2 = (\tau/\sigma)^2. \tag{20}$$

And, from (16),

$$k = \bar{X} - a\mu \tag{21}$$

These results may be expressed in the following theorem.

THEOREM I. *The estimated scale* a *in*

$$y = a^{-1}\Phi(t), \tag{8}$$

where by (7), $t = (x - k)/a$, *as obtained by the Method of Moments from observations* $X_1, X_2, \cdots, X_n$, *is the root-mean square of* $|X_i - \bar{X}|/\sigma$, *where* $\bar{X}$ *is the arithmetic mean of the* X_i*'s, and* σ^2 *is the "theoretic" variance of* $\Phi(t)$ *itself, as a function of t—with no reference to the* k *or* a *in* (7).

Moreover, the estimated location k is a substitutive mean, characterized by (3), and given by

$$k = \bar{X} - a\mu = \bar{X} - [\Sigma(X_i - \bar{X})^2/\sigma^2 n]^{\frac{1}{2}}. \tag{21}$$

As regards this final statement, it will be seen that if each $X_i = c$, then $\bar{X} = c$; and hence $k = c$,—as required by (3). We may say, then, that the right member of (21) obtained as the formal solution of equations which the problem sets up, is a substitutive mean of the elements X_i.

But if each $X_i = c$, then $a = 0$ in (21); and this a may not be used as a scale However, if any two X_i's are different, $a \neq 0$ And it is evident that as $X_i \to c$ simultaneously, limit $k = c$ If, then, we consider that the right member of (21) is not originally defined for equal values c of the elements X_i, it is to be given its "continuity" value c, in accordance with the common convention already mentioned.

In the special case where the function Φ in (5) chosen to specify the distribution has a first moment μ equal to zero, the estimate k of location given by (21) is seen at once to be the arithmetic mean of the observations $X_1, X_2, \cdots, X_n$

3 **External means.** In the papers cited, Chisini and DeFinetti gave examples of external means Indeed, it is not difficult to find means which do not conform to the condition of internality:

(22) $$\text{Minimum } (X_i) \leqq \text{Mean } (X_i) \leqq \text{Maximum } (X_i).$$

As a simple illustration, suppose that there are just three measurements $X_1 = 1$, $X_2 = 1$, $X_3 = -2$ The standard deviation $\sqrt{2}$ is greater than each measurement—it is an external mean. In this case also, the estimate of scale mentioned in Theorem I is an external mean of $(X_i - \bar{X})/\sigma$ But, it may be noted that a, the estimate of scale, is an *internal* mean of $|X_i - \bar{X}|/\sigma$

However, it will be shown now that the estimate k of location may be an external mean, with an *externality not "removable"* by the simple device of using absolute values.

And it may be noted that in the earlier paper cited, I found by the *Method of Maximum Likelihood* estimates of the *scale* a, which were likewise *not removable*

THEOREM II. *If for the function $\Phi(t)$ in (8), the second moment is less than twice the square of the second moment, then the estimated location given by*

(21) $$k = X - a\mu$$

is an external mean of the measurements X_i, if these are all numerically equal, half of them positive and the other half negative.

Proof Let the positive measurements be c, and the negative measurements be $-c$. Then $\bar{X} = 0$; also in (19), $\tau = c$ Hence from (20), $a = c/\sigma$ But by hypothesis, the second moment $\sigma^2 + \mu^2$ of $\Phi(t)$ in (11) is less than $2\mu^2$, and thus $|\mu/\sigma| > 1$. Then, by (21) $k = \bar{X} - a\mu = (-c/\sigma)\mu$; and hence $|k| > c$. Either k is greater than every positive measurement c, or it is less than every negative measurement $-c$. In either case, it is an external mean

COROLLARY *If in $\Phi(t)$, the t is subjected to a translation $t = u + b$, so that $\Phi(t) = \Phi(u + b) = \Psi(u)$, then it is always possible to choose b so that the second moment of $\Psi(u)$ is less than twice the square of its first moment; and thus if a location k' is obtained from $\Psi(u)$, external means may occur On the other hand, by proper choice of b, it is possible to make the first moment zero, so that the location becomes the arithmetic mean $\bar{X}$ of the X_i's*

The first part of this corollary may be seen from (11) which states that Second

Moment $= \mu^2 + \sigma^2$. Translation does not change σ^2, but it can increase μ^2 indefinitely,—making eventually $\mu^2 > \sigma^2$, and thus $\mu^2 + \sigma^2 < 2\mu^2$

4 **Illustration.** For the Pearson Type III the simplest specification is perhaps with the origin at the start In this case,

$$\Phi(t) = (p!)^{-1}e^{-t}t^p, \qquad p > -1, \qquad t \geqq 0, \qquad t = (x - \text{k})/\text{a}. \tag{23}$$

Here $p! = \Gamma(1 + p)$ Apart from this numerical factor, $\Phi(t)$ is the integrand of the Gamma function. With $\Phi(t)$ in this form, it is easily seen that the first moment is $(p + 1)$ and the second moment is $(p + 1)(p + 2)$ In the usual[6] case, $p > 0$. Here, then, the second moment is less than twice the square of the first moment. If, then, there are an even number of measurements, all numerically equal, with half the measurements positive and the other half negative, then the estimate k of location as found by the Method of Moments is an *external* mean of the measurements. Such conditions, while sufficient, are by *no means necessary* for externality

5. **Summary.** Suppose that the specification for a frequency function in x is $\alpha^{-1}\Phi(t')$, where $t' = (x - \kappa)/\alpha$, and that for the unknown scale α and location κ, estimates a and k, respectively, are made by the Method of Moments from a set of n measurements X_i with arithmetic mean $\bar{X}$. Let σ^2 be the variance of $\Phi(t')$. Then the estimate a is the root-mean-square of $| X_i - \bar{X} | /\sigma$, an internal mean. The estimate k of the location is $\bar{X} - \mu\text{a}$, where μ is the first moment of $\Phi(t')$ This is a substitutive mean of the measurements X_i; and it may be external—either greater than Maximum X_i or less than Minimum X_i.

THE UNIVERSITY OF TEXAS

[6] W. Palin Elderton, *Frequency Curves and Correlation*, Second Edition, p 91

ON THE CHI-SQUARE DISTRIBUTION FOR SMALL SAMPLES[1]

By Paul G. Hoel

1. Introduction. The use of what is known as the χ^2 distribution function for testing goodness of fit involves two types of error. One arises from the fact that the derivation of this function is based upon rough approximations, while the other arises from using the integral of this continuous function in place of summing the proper terms of a discrete set. Both of these errors become increasingly important as the sample becomes small. The purpose of this paper is to investigate the nature of this first type of error by finding a better approximation than the customary one to what might be called the exact continuous χ^2 distribution function.

The method employed is that of generating or characteristic functions, and consists in expressing successively in expanded form the generating function of the multinomial, the distribution function of the multinomial, the generating function of χ^2, and the distribution function of χ^2. Only the first and second order terms of this final distribution function are evaluated explicitly because of the increasingly heavy algebra involved. By means of these second order terms, the nature of the error involved in the use of the customary first order approximation is investigated.

2. The Generating Function of the Multinomial. Consider $k + 1$ cells into which observations can fall, and let p_i be the probability that an observation will fall in cell i. If n observations are made, the probability that cell i will contain α_i of these observations is given by the multinomial

$$P = \frac{n!}{\alpha_1!\,\alpha_2!\cdots\alpha_{k+1}!}\,p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_{k+1}^{\alpha_{k+1}},$$

where $\sum_{i=1}^{k+1}\alpha_i = n$. The generating function of this multinomial can be written as[2]

$$M = [p_1e^{t_1} + \cdots + p_ke^{t_k} + p]^n = \left[1 + \sum_{i=1}^{k} p_i(e^{t_i} - 1)\right]^n,$$

where α_{k+1} is chosen as the dependent variable and p_{k+1} is written as p.

[1] Presented to the American Mathematical Society, April 9, 1938.

[2] Cf. Darmois, *Statistique Mathematique*, pp. 237–242, for the methods used in this and the next two paragraphs.

Let $x_i = \dfrac{\alpha_i - np_i}{\sqrt{n}}$. The generating function of the x_i is obtained from that of the α_i by multiplying M by the proper factor to shift the origin to the mean and then replacing t_i by, say, $u_i/\sqrt{n}$ to compensate for the change in scale. Denoting this function by φ,

$$\varphi = e^{-n\sum_{i=1}^{k}\frac{p_i u_i}{\sqrt{n}}}\left[1 + \sum_{i=1}^{k} p_i(e^{u_i/\sqrt{n}} - 1)\right]^n.$$

Consequently,

$$\log \varphi = -\sqrt{n}\sum_{i=1}^{k} p_i u_i + n \log\left[1 + \sum_{i=1}^{k} p_i(e^{u_i/\sqrt{n}} - 1)\right]$$

Since the range of the u_i may be selected sufficiently small for convergence, the logarithm on the right may be expanded in powers of the summation, which in turn may be expanded in powers of the u_i. Terms containing $1/n^{\frac{1}{2}q}$ as a factor will be homogeneous in the u_i of degree $q + 2$. Writing down only the terms of order $1/n$ and lower, this double expansion gives

$$\begin{aligned}
\log \varphi = {} & \frac{1}{2}\left[\sum_{i=1}^{k} (p_i - p_i^2)u_i^2 - 2\sum_{i<j} p_i p_j u_i u_j\right] \\
& + \frac{1}{\sqrt{n}}\left[\frac{1}{6}\sum_{i=1}^{k} (p_i - 3p_i^2 + 2p_i^3)u_i^3 - \frac{1}{2}\sum_{i\neq j} (p_i p_j - 2p_i^2 p_j)u_i^2 u_j \right.\\
& \left. + 2\sum_{i<j<l} p_i p_j p_l u_i u_j u_l\right] + \frac{1}{n}\left[\frac{1}{24}\sum_{i=1}^{k} (p_i - 7p_i^2 + 12p_i^3 - 6p_i^4)u_i^4 \right.\\
& - \frac{1}{6}\sum_{i\neq j} (p_i p_j - 6p_i^2 p_j + 6p_i^3 p_j)u_i^3 u_j \\
& - \frac{1}{4}\sum_{i<j} (p_i p_j - 2p_i^2 p_j - 2p_i p_j^2 + 6p_i^2 p_j^2)u_i^2 u_j^2 \\
& + \sum_{\substack{i<j<l \\ j<i<l \\ j<l<i}} (p_i p_j p_l - 3p_i^2 p_j p_l)u_i^2 u_j u_l \\
& \left. + 6\sum_{i<j<l<m} p_i p_j p_l p_m u_i u_j u_l u_m\right] + \cdots
\end{aligned} \tag{1}$$

Hence φ can be written in the form

$$\varphi = e^{\frac{1}{2}\left[\sum_{i=1}^{k}(p_i - p_i^2)u_i^2 - 2\sum_{i<j} p_i p_j u_i u_j\right]}\left[1 + \frac{A_1}{\sqrt{n}} + \frac{A_2}{n} + \cdots\right], \tag{2}$$

where A_1 is the coefficient of $1/\sqrt{n}$ in (1), A_2 is the sum of the coefficient of $1/n$ and $A_1^2/2$, etc.

3 **The Multinomial Distribution Function.** If a distribution function can be expressed as

$$f(x_1, \cdots, x_k) = f_0 - \sum_{i=1}^{k} \alpha_i \frac{\partial f_0}{\partial x_i} + \sum_{i,j=1}^{k} \beta_{ij} \frac{\partial^2 f_0}{\partial x_i \partial x_j} - \cdots, \tag{3}$$

where f_0 is of the form $c_0 e^{-\frac{1}{2}\Sigma c_{ij} x_i x_j}$ with $|c_{ij}|$ positive definite, then its generating function[3] can be written as

$$F(u_1, \cdots, u_k) = F_0\left[1 + \sum_{i=1}^{k} \alpha_i u_i + \sum_{i,j=1}^{k} \beta_{ij} u_i u_j + \cdots\right], \tag{4}$$

where F_0 is the generating function of f_0 and is of the form $e^{\frac{1}{2}\Sigma a_{ij} u_i u_j}$ with $|a_{ij}|$ positive definite. Conversely,[4] if the generating function of a continuous distribution is of the form (4), then the distribution function can be expressed by means of (3). This relationship may be applied to (2) since it can be shown to be of the form (4).

The coefficients c_{ij} of f_0 corresponding to φ can be determined by making use of the fact that the moments of f_0 can be evaluated directly by integration or indirectly by differentiation of φ_0 It is sufficient here to equate expressions for second moments; thus

$$\left.\frac{\partial^2 \varphi_0}{\partial u_s \partial u_t}\right|_{u_i=0} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_s x_t c_0 e^{-\frac{1}{2}\Sigma c_{ij} x_i x_j} dx_1 \cdots dx_k$$

Now

$$\left.\frac{\partial^2 \varphi_0}{\partial u_s \partial u_t}\right|_{u_i=0} = \begin{cases} -p_s p_t, & s \neq t \\ p_s - p_s^2, & s = t \end{cases}$$

The value of the integral is known[5] to be c^{st}, the reciprocal of the element c_{st} in the determinant $|c_{ij}|$. Hence

$$c^{st} = \begin{cases} -p_s p_t, & s \neq t \\ p_s - p_s^2, & s = t \end{cases}$$

But c_{st} can be obtained from c^{st}, since it is given by the reciprocal of c^{st} Thus $c_{st} = \bar{c}^{st}/|c^{st}|$, where $\bar{c}^{st}$ denotes the cofactor of element c^{st} in $|c^{st}|$.

---

[3] Darmois, loc. cit., p. 242.

[4] See, for example, S Kullback, *Annals of Mathematical Statistics*, vol. 5 (1934), pp 263–307.

[5] See, for example, Risser and Traynard, *Les Principes de la Statistique Mathematique*, p. 226.

$$|c^{st}| = \begin{vmatrix} p_1 - p_1^2 & -p_1p_2 & \cdot & -p_1p_k \\ & p_2 - p_2^2 & \cdot & -p_2p_k \\ & & \ddots & \vdots \\ & & & p_k - p_k^2 \end{vmatrix}$$

$$= (-1)^k p_1^2 p_2^2 \cdots p_k^2 \begin{vmatrix} 1 - \frac{1}{p_1} & 1 & \cdots & 1 \\ & 1 - \frac{1}{p_2} & \cdots & 1 \\ & & \ddots & \vdots \\ & & & 1 - \frac{1}{p_k} \end{vmatrix}$$

This determinant may be evaluated by subtracting the last column from each of the others and then expanding by minors of the last row. Thus

$$|c^{st}| = (-1)^k p_1^2 \cdots p_k^2 \left[(-1)^k \frac{1 - \sum_{i=1}^{k} p_i}{p_1 \cdots p_k} \right] = p_1 p_2 \cdots p_k p,$$

since $\sum_{i=1}^{k} p_i = 1 - p$ from probability considerations To evaluate $\bar{c}^{st}$, delete row s and column t in $|c^{st}|$, then shift row t to the last row and column s to the last column. These shifts, together with the sign of the cofactor, change the sign of the resulting expression; hence

$$\bar{c}^{st} = -(-1)^{k-1} \frac{p_1^2 \cdots p_k^2}{p_s p_t} \begin{vmatrix} 1 - \frac{1}{p_1} & 1 & \cdots & 1 \\ & 1 - \frac{1}{p_2} & \cdots & 1 \\ & & \ddots & \vdots \\ & & 1 - \frac{1}{p_k} & 1 \\ & & & 1 \end{vmatrix} = p_1 p_2 \cdots p_k,$$

provided $s \neq t$. Since $\bar{c}^{ss}$ is merely $|c^{st}|$ after row s and column s have been deleted, it may be evaluated exactly as was $|c^{st}|$. Thus

$$\bar{c}^{ss} = (-1)^{k-1} \frac{p_1^2 \cdots p_k^2}{p_s^2} \left[(-1)^{k-1} p_s \frac{1 - \sum_{i \neq s}^{k} p_i}{p_1 \cdots p_k} \right] = p_1 p_2 \cdots p_k \left(1 + \frac{p}{p_s} \right).$$

Combining these results, $c_{st} = \frac{1}{p}$ for $s \neq t$ and $c_{ss} = \frac{1}{p} + \frac{1}{p_s}$; and therefore

$$f_0 = c_0 e^{-\frac{1}{2}\left[\sum_{i=1}^{k}\left(\frac{1}{p}+\frac{1}{p_i}\right)x_i^2+\frac{2}{p}\sum_{i<j}x_ix_j\right]} \tag{5}$$

By computing the necessary derivatives of f_0, the explicit form of f, given by (3), can be obtained to the desired number of terms. Since such derivatives contain f_0 as a factor, f may be written as

$$f = f_0\left[1 - \frac{B_1}{\sqrt{n}} + \frac{B_2}{n} - \cdots\right], \tag{6}$$

where B_i is obtained from A_i of (2) by replacing terms in the u_i with the corresponding derivative of f_0 and then factoring out f_0.

4. **The Generating Function of χ^2.** Let this function be denoted by

$$G(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{t\chi^2} f_0\left[1 - \frac{B_1}{\sqrt{n}} + \frac{B_2}{n} - \cdots\right] dx_1 \cdots dx_k.$$

Now

$$\chi^2 = \sum_{i=1}^{k+1}\left(\frac{a_i - np_i}{\sqrt{np_i}}\right)^2 = \sum_{i=1}^{k+1}\frac{x_i^2}{p_i} = \sum_{i=1}^{k}\left(\frac{1}{p}+\frac{1}{p_i}\right)x_i^2 + \frac{2}{p}\sum_{i<j}x_ix_j;$$

consequently χ^2 is, except for a factor of $-\frac{1}{2}$, the quadratic form in f_0. Accordingly, letting $\theta = 1 - 2t$,

$$e^{t\chi^2} f_0 = e^{t\chi^2} c_0 e^{-\frac{1}{2}\chi^2} = c_0 e^{-\frac{1}{2}\theta\chi^2};$$

and hence

$$G(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} c_0 e^{-\frac{1}{2}\theta\chi^2}\left[1 - \frac{B_1}{\sqrt{n}} + \frac{B_2}{n} - \cdots\right] dx_1 \cdots dx_k.$$

Letting $z_i = x_i\sqrt{\theta}$ and denoting the value of B_i after this substitution by C_i,

$$\begin{aligned} G(t) &= \theta^{-\frac{1}{2}k} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_0\left[1 - \frac{C_1}{\sqrt{n}} + \frac{C_2}{n} - \cdots\right] dz_1 \cdots dz_k, \\ &= \theta^{-\frac{1}{2}k}\left[1 + \frac{1}{n}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_0 C_2\, dz_1 \cdots dz_k + \cdots\right], \end{aligned} \tag{7}$$

since the terms involving odd powers of $1/\sqrt{n}$ are of odd degree in the z_i and therefore vanish upon integration.

For the purposes of this paper only the integral which is the coefficient of $1/n$ needs to be evaluated. Since the algebra involved in this evaluation is heavy and the formulas become exceedingly long, only a few terms will be written out explicitly to indicate the procedure followed.

From (1), (2), and (6) it is clear that only fourth and sixth order derivatives of f_0 are needed As examples

$$\frac{\partial^4 f_0}{\partial x_i^4} = f_0\left[D_i^4 - 6D_i^2\left(\frac{1}{p}+\frac{1}{p_i}\right) + 3\left(\frac{1}{p}+\frac{1}{p_i}\right)^2\right],$$

$$\frac{\partial^6 f_0}{\partial x_i^6} = f_0\left[D_i^6 - 15D_i^4\left(\frac{1}{p}+\frac{1}{p_i}\right) + 45D_i^2\left(\frac{1}{p}+\frac{1}{p_i}\right)^2 - 15\left(\frac{1}{p}+\frac{1}{p_i}\right)^3\right],$$

where $D_i = \dfrac{1}{p}\left[x_1 + \cdots + \left(1+\dfrac{p}{p_i}\right)x_i + \cdots + x_k\right]$ Following the procedure indicated in (6) and (7), this integral becomes

$$\begin{aligned}
\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_0\Bigg\{&\frac{1}{24}\sum_{i=1}^{k}[p_i - 7p_i^2 + 12p_i^3 - 6p_i^4] \\
&\left[\frac{D_i^4}{\theta^2} - 6\frac{D_i^2}{\theta}\left(\frac{1}{p}+\frac{1}{p_i}\right) + 3\left(\frac{1}{p}+\frac{1}{p_i}\right)^2\right] \\
&+ \text{(similar terms of degree 4 and lower in the } D_i) \\
&+ \frac{1}{72}\sum_{i=1}^{k}[p_i^2 - 6p_i^3 + 13p_i^4 - 12p_i^5 + 4p_i^6] \\
&\left[\frac{D_i^6}{\theta^3} - 15\frac{D_i^4}{\theta^2}\left(\frac{1}{p}+\frac{1}{p_i}\right) + 45\frac{D_i^2}{\theta}\left(\frac{1}{p}+\frac{1}{p_i}\right)^2 - 15\left(\frac{1}{p}+\frac{1}{p_i}\right)^3\right] \\
&+ \text{(similar terms of degree 6 and lower in the } D_i)\Bigg\}\, dz_1 \cdots dz_k
\end{aligned} \tag{8}$$

When $\theta = 1$, the integral reduces to that of $f_0 B_3$, which in turn is the integral of a linear combination of derivatives of f_0 But the integral of such a derivative vanishes. As a result, if the integral of $f_0 D_i^2$ has been computed directly, that of $f_0 D_i^4$ and then that of $f_0 D_i^6$ can be found indirectly by equating the corresponding bracket to zero for $\theta = 1$. Similarly for the other terms of the above integral. As examples

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_0 D_i^2\, dz_1 \cdots dz_k = \frac{1}{p}+\frac{1}{p_i},$$

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_0 D_i^4\, dz_1 \cdots dz_k = 3\left(\frac{1}{p}+\frac{1}{p_i}\right)^2.$$

Upon evaluating all such integrals, (8) reduces to

$$\begin{aligned}
&\frac{1}{24}\sum_{i=1}^{k}(p_i - 7p_i^2 + 12p_i^3 - 6p_i^4)\, 3\left(\frac{1}{p}+\frac{1}{p_i}\right)^2\left(\frac{1}{\theta}-1\right)^2 \\
&\qquad + \left(\text{similar terms all containing } \left(\frac{1}{\theta}-1\right)^2 \text{ as a factor}\right) \\
&\qquad + \frac{1}{72}\sum_{i=1}^{k}(p_i^2 - 6p_i^3 + 13p_i^4 - 12p_i^5 + 4p_i^6)\, 15\left(\frac{1}{p}+\frac{1}{p_i}\right)^3\left(\frac{1}{\theta}-1\right)^3 \\
&\qquad + \left(\text{similar terms all containing } \left(\frac{1}{\theta}-1\right)^3 \text{ as a factor}\right)
\end{aligned} \tag{9}$$

In order to interpret these results, it is necessary to condense these various sums of probability expressions. If the terms are arranged in descending powers of the p_i, it will be discovered that certain combinations condense readily. The condensation in each case lies in recognizing combinations like

$$\sum_{i=1}^{k} p_i^4 + 4\sum_{i\neq j} p_i^3 p_j + 6\sum_{i<j} p_i^2 p_j^2 + 12\sum_{i<j<l} p_i^2 p_j p_l$$
$$+ 24\sum_{i<j<l<m} p_i p_j p_l p_m = \left(\sum_{i=1}^{k} p_i\right)^4.$$

However, some of the terms resulting from multiplying by $1/p_i$ above cannot be condensed in this fashion until they have been reduced to familiar sums by using relationships of the following type:

$$\sum_{i<j} p_i p_j\left(\frac{1}{p_i} + \frac{1}{p_j}\right) = (k-1)\sum_{i=1}^{k} p_i,$$

$$\sum_{i<j<l} p_i p_j p_l\left(\frac{1}{p_i} + \frac{1}{p_j} + \frac{1}{p_l}\right) = (k-2)\sum_{i<j} p_i p_j$$

After all possible condensations have been made, (9) reduces to

$$\left(\frac{1}{\theta} - 1\right)^2 \frac{1}{8}\left[\sum_{i=1}^{k+1}\frac{1}{p_i} - (k^2 + 4k + 1)\right]$$
$$+ \left(\frac{1}{\theta} - 1\right)^3 \frac{1}{24}\left[5\sum_{i=1}^{k+1}\frac{1}{p_i} - (3k^2 + 12k + 5)\right].$$

As a result, the generating function of χ^2 can be written as

$$G(t) = \theta^{-\frac{1}{2}k} + \frac{S_1}{n}\left(\theta^{-\frac{1}{2}(k+4)} - 2\theta^{-\frac{1}{2}(k+2)} + \theta^{-\frac{1}{2}k}\right)$$

$$(10)\qquad + \frac{S_2}{n}\left(\theta^{-\frac{1}{2}(k+6)} - 3\theta^{-\frac{1}{2}(k+4)} + 3\theta^{-\frac{1}{2}(k+2)} - \theta^{-\frac{1}{2}k}\right)$$

$$+ \text{(terms involving higher powers of } 1/n),$$

where $S_1 = \frac{1}{8}\left[\sum_{i=1}^{k+1}\frac{1}{p_i} - (k^2 + 4k + 1)\right]$ and $S_2 = \frac{1}{24}\left[5\sum_{i=1}^{k+1}\frac{1}{p_i} - (3k^2 + 12k + 5)\right]$.

5. The Distribution Function of χ^2 It is well known that $\theta^{-\frac{1}{2}k} = (1 - 2it)^{-\frac{1}{2}k}$ is the generating function of what is commonly called the χ^2 distribution function with k degrees of freedom. If this distribution function is denoted by $F_k(\chi^2)$, then the distribution function corresponding to (10) can be written as

$$(11)\qquad F_k + \frac{S_1}{n}(F_{k+4} - 2F_{k+2} + F_k) + \frac{S_2}{n}(F_{k+6} - 3F_{k+4} + 3F_{k+2} - F_k)$$

$$+ \text{(terms involving higher powers of } 1/n).$$

The customary test for goodness of fit involves the integral of $F_k(\chi^2)$ from χ to ∞, which has been tabled for values of χ^2 and k. The form of (11) is such that the integral of the term in $1/n$ is easily evaluated by means of this same table. However, for more accurate results and for theoretical reasons, it is more elucidating to express these integrals in a more compact form. This is accomplished by using familiar[6] expansions for the integral of $F_k(\chi^2)$. Denoting the integral of the explicit terms of (11) by P, it is easy to show that

$$P = P_1 + \frac{1}{n}[R_1S_1 + R_2S_2], \tag{12}$$

where P_1 is the customary tabled value for k degrees of freedom and

$$\begin{aligned} R_1 &= \frac{e^{-\frac{1}{2}\chi^2}\chi^k}{2.4 \cdots (k+2)}[\chi^2 - (k+2)], \\ R_2 &= \frac{e^{-\frac{1}{2}\chi^2}\chi^k}{2.4 \cdots (k+4)}[\chi^4 - 2(k+4)\chi^2 + (k+4)(k+2)], \end{aligned} \tag{13}$$

for k even, while for k odd both R_1 and R_2 contain an additional factor of $\sqrt{2/\pi}$ and have $1.3 \cdots (k+2)$ and $1.3 \cdots (k+4)$ respectively for denominators.

6. Conclusions. In any given problem the second approximation P can be calculated easily by means of either (11) or (12) and compared with the customary first approximation P_1. However, the magnitude of this correction term is of primary interest when χ^2 is near a significance level and when one or more of n, k, and p_i is small because the accuracy of P_1 is questioned in those cases

For χ^2 at the .05 level and for $2 \leq k \leq 16$, it is easily shown that $0 < R_1 < .08$ and $-.08 < R_2 < 0$ Clearly S_2 is positive, while S_1 will be positive if one or more of the p_i is sufficiently small Consequently, for those cases of particular interest, the correction term is surprisingly small partly because R_1 and R_2 are so small and partly because they are of opposite sign

To illustrate this viewpoint consider the following numerical example Let $n = 10$, $k = 4$, $\chi^2 = 9.488$, $p_1 = p_2 = \frac{1}{20}$, $p_3 = \frac{4}{20}$, $p_4 = \frac{6}{20}$, $p_5 = \frac{8}{20}$ Then $S_1 = 2.23$, $S_2 = 6.38$, $R_1 = .056$, $R_2 = -.027$, $P_1 = .05$, and $P = .045$ The correction term of $-.005$ is very small in spite of the fact that this example is an extreme case to which the customary χ^2 test would not be applied.

As judged by the second order approximation obtained in this paper, the actual error comitted by using the customary first approximation is much smaller than the order of the neglected terms would indicate, and therefore the range of applicability of P_1 is wider than has been supposed However, this investigation has considered only the error due to rough approximations and leaves untouched the second type of error indicated in the introductory paragraph

OREGON STATE COLLEGE

[6] Risser and Traynard, loc. cit., p. 251

SHORTEST AVERAGE CONFIDENCE INTERVALS FROM LARGE SAMPLES

By S S Wilks

1 **Introduction.** The method of fiducial argument [1, 2] in statistics has gained considerable prominence within the last few years as a method of inferring the values of population parameters from samples "randomly drawn" from populations having distribution laws of known functional forms. The method has also been shown to be applicable [2] to the problem of inferring the values of statistical functions in samples from samples already observed, assuming all samples to be drawn from a population with a distribution law of a given functional form

The main ideas of a procedure which is sufficient for carrying out fiducial inference for certain cases of a single population parameter may be summed up in the following steps:

(a) A sample is assumed to be "randomly drawn" from a population with a distribution law $f(x, \theta)$ of known functional form

(b) A function $\psi(x_1, x_2, \cdots, x_n, \theta)$ of the sample values $x_1, x_2, \cdots, x_n$ and parameter θ is devised, which is a monotonic function of θ for a given sample, so that the sampling distribution $G(\psi)d\psi$ of $\psi(x_1, x_2, \cdots, x_n, \theta_0) = \psi_0$, say, in samples from the population with $\theta = \theta_0$ is independent of θ_0 and the x's, except as they enter into ψ

(c) For a given probability α a pair of numbers ψ'_α and ψ''_α is chosen so that when $\theta = \theta_0$, the probability that $\psi'_\alpha < \psi_0 < \psi''_\alpha$ is $1 - \alpha$, or more, briefly,

$$P(\psi'_\alpha < \psi_0 < \psi''_\alpha \mid \theta = \theta_0) = 1 - \alpha \tag{1}$$

which can be stated in the alternative form

$$P(\underline{\theta} < \theta_0 < \bar{\theta} \mid \theta = \theta_0) = 1 - \alpha. \tag{2}$$

(d) $\underline{\theta}$ and $\bar{\theta}$ being functions of ψ'_α, ψ''_α and the sample, are subject to sampling fluctuations and it can be stated that the probability is $1 - \alpha$ that they will include the true value of θ, whatever it may be, that is, θ_0, between them. The statement holds for all values which θ_0 may take on.

The numbers $\underline{\theta}$ and $\bar{\theta}$ are known as *fiducial* or *confidence limits* [3] of θ_0 and $(\underline{\theta}, \bar{\theta})$ a *confidence interval* for the *confidence coefficient* $1 - \alpha$. We therefore have the following rule for making inferences about the unknown number θ_0 once ψ has been found: For a given sample solve the equations

$$\psi(x_1, x_2, \cdots, x_n, \theta_0) = \psi'_\alpha, \qquad \psi(x_1, x_2, \cdots, x_n, \theta_0) = \psi''_\alpha$$

for θ_0 Let $\underline{\theta}$ and $\bar{\theta}$ be the two values of θ_0 formally obtained. The statement that $\underline{\theta}$ and $\bar{\theta}$ will include the value of θ in the population actually sampled, if consistently made in each of an aggregate of cases involving populations having distributions of the same functional form $f(x, \theta)$ will be correct (theoretically) in $100(1 - \alpha)$ per cent of the cases.

If ψ is a function of statistics t_1 and t_2 of two samples from a population of known functional form, which is monotonic in each t for given values of the other, then one can argue fiducially about values of one t from values of the other one.

For a finite value of n and discrete distributions $f(x, \theta)$, it is not possible to carry through steps (b), (c), (d) as they are now stated. However, under certain conditions, it is possible to carry out a procedure for the discrete case which will allow one to say

$$P(\underline{\theta} < \theta_0 < \bar{\theta} \mid \theta = \theta_0) \geq 1 - \alpha. \tag{3}$$

ψ functions which have the property that their sampling distributions are independent of θ and the x's for a given distribution $f(x, \theta)$ are not, in general, unique. The question then arises as to how (if possible) one can choose ψ functions and limits ψ'_α and ψ''_α so as to get confidence intervals for a given α, which are shortest or "best" in some sense. Neyman [4] has investigated the problem of obtaining "best" confidence intervals for the case of small samples The object of this paper is to consider the problem for large samples Under fairly general conditions it will be seen that a rather simple asymptotic solution exists for the large-sample case, which is connected in an essential manner with the *method of maximum likelihood.*

2. An asymptotic distribution. Suppose a population Π has a distribution function $f(x, \theta)$, where x is a random variable and θ a parameter Actually, $f(x, \theta)$ may involve several other parameters whose values may be regarded as fixed throughout the paper. The problem of arguing fiducially about several parameters simultaneously will not be considered in this paper. In order to include the case of a discrete as well as a continuous variate x, we shall consider the cumulative distribution function (c d.f.) $F(x, \theta)$, which is monotonic and is such that

$$F(-\infty, \theta) = 0, \qquad F(+\infty, \theta) = 1, \qquad F(x + c, \theta) \geq F(x, \theta)$$

$$F(x + 0, \theta) = F(x, \theta), \text{ for } c > 0, \text{ and } a < \theta < b.$$

Thus, $F(x', \theta) = P(x \leq x' \mid \theta)$. In the case of a continuous variate x, where $f(x, \theta)$ is a probability density function, then $dF(x, \theta) = f(x, \theta)\, dx$; in the discrete case $dF(x, \theta) = f(x, \theta)$ where $f(x, \theta)$ is the probability that variate x takes the value indicated. We shall be interested in continuous functions $\varphi(x)$ for which the integral $\int \varphi(x)\, dF(x, \theta)$ taken in the Stieltjes sense exists. Limits on integral signs are understood to be $-\infty$ and ∞.

Now consider a sample O_n of n individuals independently drawn from Π_0, the population for which the c.d.f is $F(x, \theta_0)$. Let the values of x in the sample be $x_1, x_2, \cdots, x_n$. The probability element associated with the sample is

$$dP_n = \prod_{i=1}^{n} dF(x_i, \theta_0). \tag{4}$$

Let $L = \log dP_n$. Then assuming that $\frac{\partial}{\partial\theta} \log d\,F(x, \theta) = g(x, \theta)$, say, exists for $\theta = \theta_0$, and for each x, (except for a set of probability 0), we have

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^{n} g(x_i, \theta). \tag{5}$$

In all ordinary problems in statistics $g(x, \theta)$ reduces to $\frac{\partial f(x, \theta)}{\partial \theta} \Big/ f(x, \theta)$ where $f(x, \theta)$ is probability in the case where x is a discrete random variable and probability density in the case of a continuous random variable. Let g_0 denote $g(x, \theta_0)$ and $\left(\frac{\partial L}{\partial \theta}\right)_0$ denote $\frac{\partial L}{\partial \theta}$ with $\theta = \theta_0$. Let

$$A_0^2 = E_0[(g_0)^2] = \int g_0^2\, dF(x, \theta_0). \tag{6}$$

$E_0(\varphi)$ will be used to denote the mathematical expectation of φ in samples from Π_0, i.e. when the population distribution is $dF(x, \theta_0)$. We shall consider the sampling theory of

$$\psi_0 = \frac{\left(\frac{\partial L}{\partial \theta}\right)_0}{\sqrt{n}\, A_0} \tag{7}$$

in large samples, from Π_0

Let $\varphi_0^{(n)}(t)$ be the characteristic function of ψ_0 for samples from Π_0; it is defined by $E_0(e^{it\psi_0})$. Then

$$\begin{aligned} \varphi_0^{(n)}(t) &= \left\{E_0\left[\exp\left(\frac{itg_0}{\sqrt{n}\, A_0}\right)\right]\right\}^n \\ &= \left\{E_0\left[1 + \frac{itg_0}{\sqrt{n}\, A_0} - \frac{t^2 g_0^2}{2nA_0^2} + \frac{1}{n^{\frac{3}{2}}} g_0^3 (\phi_1 + i\phi_2)\right]\right\}^n \end{aligned} \tag{8}$$

where ϕ_1 and ϕ_2 are real functions of t, x, n and θ_0, such that if $|E_0(g_0)^3| \leq K < \infty$, (i.e. the third moment of g_0 is finite when $\theta = \theta_0$), for $a < \theta_0 < b$, then $E_0[g_0^3\phi_1]$ and $E_0[g_0^3\phi_2]$ are uniformly bounded for some t-interval δ(which includes $t = 0$ as an interior point) for n larger than some n_0 and for θ_0 on any fixed subinterval of the interval (a, b). Suppose that $F(x, \theta)$ is such that

$$\int \frac{\partial}{\partial \theta} dF(x, \theta) = \frac{\partial}{\partial \theta} \int dF(x, \theta) = 0, \qquad a < \theta < b \tag{9}$$

This condition implies that the range of x be independent of θ

If n is allowed to increase indefinitely, then we have at once that $\varphi_0^{(n)}(t)$ tends to $e^{-\frac{1}{2}t^2}$ uniformly in the interval δ. We now make use of a theorem [5] which states that if an unlimited sequence of random variables $x^{(1)}, x^{(2)}, \cdots, x^{(n)} \cdots$ with c.d.f.'s $F^{(1)}(x), F^{(2)}(x), \cdots, F^{(n)}(x) \cdots$ have corresponding characteristic functions $\varphi^{(1)}(t), \varphi^{(2)}(t) \cdots \varphi^{(n)}(t) \cdots$ then a necessary and sufficient condition for $F^{(n)}(x)$ to converge uniformly to a c.d.f. $F(x)$ at each point of continuity of $F(x)$ on the interval $(-\infty, \infty)$ is that the sequence of characteristic functions converge uniformly to a function $\varphi_1(t)$ on an interval $|t| < \epsilon$ for some $\epsilon > 0$. The characteristic function $\varphi(t)$ associated with $F(x)$ will then be identical with $\varphi_1(t)$ and the sequence $\varphi^{(1)}(t), \cdots \varphi^{(n)}(t) \cdots$ converges to $\varphi(t)$ uniformly in every finite t-interval.

From this theorem it follows at once that, since $e^{-\frac{1}{2}t^2}$ is the characteristic function of a variate distributed normally with mean 0 and variance 1, the asymptotic c.d.f of ψ_0 for large samples is given by

$$F(\psi_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\psi_0} e^{-\frac{1}{2}x^2}\, dx \tag{10}$$

We may conveniently summarize the foregoing results in the following

THEOREM 1. *Let $x_1, x_2, \cdots, x_n$ be the values of x in a sample of independently drawn individuals from a population Π_0 which has a c.d.f. $F(x, \theta_0)$, such that* for $a < \theta_0 < b$,

(a) $\frac{\partial}{\partial\theta} dF(x, \theta)$ *exists for all x's except possibly for a set of probability* 0;

(b) $E_0[(g_0^2)]$ *is finite; for $n > n_0$,*

(c) *condition* (9) *is satisfied.*

Then the asymptotic c.d.f. of ψ_0 for large samples defined in (7) *is given by* (10).

The statistical significance of this Theorem is that if we know the functional form $f(x, \theta)$ (for which the first derivative $f'(x, \theta)$ with respect to θ exists) of the distribution function of a population Π and if the sample $x_1, x_2, \cdots, x_n$ is "randomly drawn" from Π_0, then the quantity

$$\psi_0 = \frac{\sum_1^n \frac{f'(x_i, \theta_0)}{f(x_i, \theta_0)}}{\sqrt{n}\sqrt{E_0\left[\left(\frac{f'(x, \theta_0)}{f(x, \theta_0)}\right)^2\right]}} \tag{11}$$

is a random variable which is approximately normally distributed with mean 0 and variance 1 in repeated large samples. It will be noticed that the quantity in the numerator of (11), is simply the derivative with respect to θ, at $\theta = \theta_0$, of the logarithm of the likelihood of θ for the given sample. ψ_0 is a function of the sample O_n and the true value θ_0 of the parameter θ, and the thing that makes ψ_0 a random variable is the random nature of the sample; θ_0 is a fixed but unknown number. Thus, for example, when $1 - \alpha = .95$ in (1) and knowing that we have "randomly drawn" a large sample O_n from a population Π_0 with

distribution $f(x, \theta_0)$ of known functional form, we can say that the probability is .95 that the sample will produce a value of ψ_0 in the interval -1.96 to $+1.96$ that is,

$$P(-1.96 < \psi_0 < 1.96 \mid \theta = \theta_0) = .95 \tag{12}$$

This statement holds, whatever may be the value of the unknown θ_0. Now, the inequality $-1.96 < \psi_0 < 1.96$ is equivalent to the inequality, $\underline{\theta} < \theta_0 < \bar{\theta}$ because of the monotonic nature of ψ_0 as a function of θ_0. Hence (12) is equivalent to

$$P(\underline{\theta} < \theta_0 < \bar{\theta} \mid \theta = \theta_0) = .95 \tag{13}$$

where $\underline{\theta}$ and $\bar{\theta}$ are obtained by solving $\psi_0 = \pm 1.96$ for θ_0. The *fiducial limits* $\underline{\theta}$ and $\bar{\theta}$ will thus be functions of the sample and will be subject to sampling variations. In general, of course, one could choose any probability level $1 - \alpha$, and find ψ_α so that

$$P(-\psi_\alpha < \psi_0 < \psi_\alpha \mid \theta = \theta_0) = 1 - \alpha, \tag{14}$$

from which fiducial limits for θ_0 can be found as before.

The extension of Theorem 1 to the case in which the distribution function of the population Π involves several parameters $\theta_1, \theta_2, \cdots \theta_h$ having values in some region R of the space of θ's, is immediate. Π_0 in this case would be specified by the values $\theta_{10}, \theta_{20}, \cdots \theta_{h0}$. In fact, we can state the situation as

THEOREM 1': *Let $F(x, \theta_1, \theta_2, \cdots, \theta_h)$ denote the c.d.f. of x and (allowing i, j, k to take on values $1, 2, \cdots, h$) let*

$$\psi_{i0} = \frac{1}{\sqrt{n}}\left(\frac{\partial L}{\partial \theta_i}\right)_0 \quad \textit{where} \quad L = \sum_{j=1}^{n} \log dF(x_j, \theta_{10}, \theta_{20}, \cdots, \theta_{h0}),$$

$$g_i = \frac{\partial}{\partial \theta_i}[\log dF(x, \theta_1, \cdots, \theta_h)],$$

$$A_{ij} = E_0[g_{i0} g_{j0}] \textit{ where } g_{i0} = g_i \textit{ with } \theta_i = \theta_{i0}.$$

If, in R,

(a) $\frac{\partial}{\partial \theta_i} dF(x, \theta_1, \cdots, \theta_h)$ *exists for all x's except possibly for a set of probability* 0;

(b) $E_0(g_{i0} g_{j0} g_{k0})$ *are all finite;*

(c) $\frac{\partial}{\partial \theta_i} \int dF(x, \theta_1, \theta_2, \cdots, \theta_h) = \int \frac{\partial}{\partial \theta_i} dF(x, \theta_1, \theta_2, \cdots, \theta_h) = 0$

(d) $\| A_{ij} \|$ *is non-singular;*

then the asymptotic distribution of the ψ_{i0} in large samples from Π_0 is a normal multivariate distribution with matrix $\| A_{ij} \|$ of variances and covariances, and zero means.

A similar theorem holds for the case in which Π is a multivariate population in addition to having several parameters.

The question now arises: In what sense is the *confidence interval* between $\underline{\theta}$ and

θ as determined from ψ_0 "best"? It will be shown that the average rate of change of ψ_0 with respect to θ at $\theta = \theta_0$ is greater than that for a rather broad class of functions of the ψ type, that is functions of the observations and θ which are asymptotically normally distributed. Since we are dealing with large samples, we are only interested in values of θ in the neighborhood of θ_0, for which ψ as a function of θ is approximately linear, and demonstrating the property just stated regarding the average rate of change of ψ with respect to θ at $\theta = \theta_0$ is equivalent to showing that the two "values" of θ_0 for which $\psi_0 = \pm\psi_\alpha$, will, on the average be closer together than those computed from any other ψ function than ψ_0 of the class of functions to be considered. This class of functions will be designated as belonging to *class* C, and will now be more accurately defined.

3. Functions of class C and their asymptotic distributions. Following an argument similar to that used in proving Theorem 1, we can readily prove

THEOREM 2: *Let $h(x, \theta)$ be a function in which x has the c.d.f. $F(x, \theta)$, and which satisfies the following conditions for $a < \theta_0 < b$:*

(15) (a) $E_0[h(x, \theta_0)] = 0$;

(b) $E_0[\{h(x, \theta_0)\}^2]$ *is finite*, for $n > n_0$.

Let

$$A_0^* = E_0[\{h(x, \theta_0)\}^2] \tag{16}$$

and for a sample of values $x_1, x_2, \cdots, x_n$ let

$$\psi_0^* = \frac{\sum_1^n h(x_i, \theta_0)}{\sqrt{n}\,A_0^*} \tag{17}$$

Then the asymptotic c.d.f. of ψ_0^* for large samples from Π_0 is given by

$$F(\psi_0^*) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\psi_0^*} e^{-\frac{1}{2}x^2}\,dx.$$

We shall designate as belonging to class C any function ψ_0^* made up according to the rule expressed by (17), of functions $h(x, \theta)$ satisfying (a) and (b) in (15) and such that ψ_0^* is asymptotically normally distributed with zero mean and unit number. Clearly, ψ_0 as defined by (7) belongs to class C.

4. Comparison of average confidence intervals computed from ψ_0 and ψ_0^*. We shall now show that for each fixed value θ_0 of θ the average rate of change of ψ with respect to θ is greater than that of ψ^* for any $h(x, \theta)$ which is not a constant multiple of $g(x, \theta)$. Consider $\frac{\partial\psi}{\partial\theta}$ and $\frac{\partial\psi^*}{\partial\theta}$ for a given n. We have,

$$\frac{\partial\psi}{\partial\theta} = \frac{1}{\sqrt{n}A}\left\{\sum_1^n \frac{\partial g(x_i, \theta)}{\partial\theta} - \frac{1}{A}\sum_1^n g(x_i, \theta)\frac{\partial A}{\partial\theta}\right\} \tag{18}$$

$$\frac{\partial\psi^*}{\partial\theta} = \frac{1}{\sqrt{n}A^*}\left\{\sum_1^n \frac{\partial h(x_i, \theta)}{\partial\theta} - \frac{1}{A^*}\sum_1^n h(x_i, \theta)\frac{\partial A^*}{\partial\theta}\right\} \tag{19}$$

Now

$$\frac{\partial g(x_i, \theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left\{\frac{\partial[dF(x_i, \theta)]}{\partial \theta} \bigg/ dF(x_i, \theta)\right\} = \frac{\frac{\partial^2}{\partial \theta^2}[dF(x_i, \theta)]}{dF(x_i, \theta)} - [g(x_i, \theta)]^2$$

Assuming that

$$\int \frac{\partial^2}{\partial \theta^2} dF(x, \theta) = \frac{\partial^2}{\partial \theta^2} \int dF(x, \theta) = 0 \tag{20}$$

and remembering that $E_0[g(x_i, \theta_0)] = 0$, we have

$$E_0\left[\left(\frac{\partial \psi}{\partial \theta}\right)_0\right] = -\sqrt{n}\, A_0 = \Delta_1 \tag{21}$$

and

$$E_0\left[\left(\frac{\partial \psi^*}{\partial \theta}\right)_0\right] = \frac{\sqrt{n}}{A_0^*} E_0\left[\left(\frac{\partial h(x, \theta)}{\partial \theta}\right)_0\right] = \Delta_2. \tag{22}$$

Now, since

$$\int h(x, \theta)\, dF(x, \theta) = 0 \tag{23}$$

and assuming that (23) can be differentiated under the integral sign, we have

$$E_0\left[\left(\frac{\partial h(x, \theta)}{\partial \theta}\right)_0\right] = \left[-\int h(x, \theta) \frac{\partial}{\partial \theta} dF(x, \theta)\right]_{\theta=\theta_0}. \tag{24}$$

For the difference $\Delta_1^2 - \Delta_2^2$ in samples from populations with $\theta = \theta_0$, we have

$$\frac{n}{(A_0^*)^2}\left\{\int \left(\frac{\frac{\partial}{\partial \theta} dF(x, \theta)}{dF(x, \theta)}\right)^2 dF(x, \theta) \cdot \int (h(x, \theta))^2\, dF(x, \theta) - \left[\int (h(x, \theta) \sqrt{dF(x, \theta)}) \left(\frac{\frac{\partial}{\partial \theta} dF(x, \theta)}{\sqrt{dF(x, \theta)}}\right)\right]^2\right\}_{\theta=\theta_0}. \tag{25}$$

Making use of Schwartz' inequality which states that

$$\int g^2(x)\, dx \cdot \int h^2(x)\, dx \geq \left[\int g(x)\, h(x)\, dx\right]^2,$$

where the equality sign holds only if $g(x) \equiv K\, h(x)$, K being a constant, it is evident that independently of n, $\Delta_1^2 \geq \Delta_2^2$, and furthermore, the only condition under which $\Delta_1^2 = \Delta_2^2$ is that

$$h(x, \theta) \sqrt{dF(x, \theta)} \equiv K \frac{\frac{\partial}{\partial \theta} dF(x, \theta)}{\sqrt{dF(x, \theta)}},$$

that is,

$$h(x, \theta) \equiv K\, g(x, \theta). \tag{26}$$

Therefore we have

THEOREM 3. *If $g(x, \theta)$ and $h(x, \theta)$ satisfy the conditions of Theorems 1 and 2 respectively and furthermore, if (20) is satisfied and if the expression on the left in (23) can be differentiated under the integral sign with respect to θ, then the average rate of change of ψ with respect to θ for each fixed value θ_0 of θ is greater than that of ψ^* (for which $h(x, \theta) \not\equiv Kg(x, \theta)$ with respect to θ, when $\theta = \theta_0$ in samples from Π_0.*

This Theorem simply means that when computed from ψ_0 the fiducial limits for the true but unknown value θ_0 of the parameter θ, whatever value θ_0 may have on the interval $a < \theta_0 < b$ of possible values, are (for large samples) closer together *on the average* than those computed from any other ψ_0^* of class C. There is no function ψ_0^* which is more *efficient*, as it were, for determining confidence intervals for θ_0 than the particular ψ_0 given by (7) which is ψ_0^* with $h(x, \theta)$ replaced by $g(x, \theta)$, that is, $\frac{\partial}{\partial\theta} \log dF(x, \theta)$. The actual manner in which the fiducial limits for θ_0 are found for a given confidence coefficient $1 - \alpha$, is to set

$$\frac{\left(\dfrac{\partial L}{\partial \theta}\right)_0}{\sqrt{n}\, A_0} = \pm \psi_\alpha \tag{27}$$

and solve formally, for θ_0, where ψ_α is the value for which $\frac{2}{\sqrt{2\pi}} \int_{\psi_\alpha}^{\infty} e^{-\frac{1}{2}x^2}\, dx = \alpha$, which can be found from normal probability tables. The two values of θ_0 thus found are the fiducial limits $\underline{\theta}$ and $\bar{\theta}$ for the true value θ_0 and we can state that the probability is $1 - \alpha$ that $\underline{\theta}$ and $\bar{\theta}$ will include the true value θ_0 between them. This statement is valid whatever may be the value of θ_0 between a and b. This rule consistently followed for large samples, will produce fiducial limits $\underline{\theta}$ and $\bar{\theta}$ which are closest together *on the average*, for *each fixed value* of the probability level α between 0 and 1. It should be observed that no assumptions have been made regarding the existence of sufficient statistics.

5. Examples. EXAMPLE 1. Suppose a large sample of n individuals to be drawn from a population known to have the Poisson distribution law

$$f(x, m) = \frac{m^x e^{-m}}{x!}.$$

We have

$$L = -\log\left(\prod_1^n x_i!\right) + (\Sigma\, x_i) \log m - nm$$

$$\left(\frac{\partial L}{\partial m}\right)_0 = \frac{\Sigma\, x_i}{m_0} - n$$

$$A_0^2 = E_0\left[\left(\frac{\partial \log f}{\partial m}\right)_0^2\right] = E_0\left[\left(\frac{x}{m_0} - 1\right)^2\right] = \frac{1}{m_0}.$$

The fiducial limits $\underline{m}$ and $\bar{m}$ for $\alpha = .05$, that is, the 95 per cent fiducial limits, are found by formally solving the equations

$$\frac{\left(\frac{\partial L}{\partial m}\right)_0}{\sqrt{n}A_0} = \frac{\left(\frac{\Sigma x_i}{m_0} - n\right)\sqrt{m_0}}{\sqrt{n}} = \pm 1.96$$

for m_0. The fiducial limits are found to be

$$\bar{x} + \frac{1.92}{n} \pm \sqrt{\frac{3.82}{n}\bar{x} + \frac{3.69}{n^2}}$$

EXAMPLE 2. Consider a large sample of n individuals known to be from a binomial population having the two classes A and B. Let p denote the probability of an individual's belonging to A, and $q = 1 - p$ that of belonging to B. Let x denote the number of individuals belonging to A in one drawing from the population; x will take on only two possible values, 1 and 0, with probabilities p and q respectively. The population distribution is thus

$$f(x, p) = p^x(1 - p)^{1-x}.$$

We have

$$L = (\Sigma x_i) \log p + \Sigma(1 - x_i) \log (1 - p)$$

$$\left(\frac{\partial L}{\partial p}\right)_0 = \frac{m}{p_0} - \frac{n - m}{1 - p_0} = \frac{m - np_0}{p_0(1 - p_0)}$$

where m is the number of individuals belonging to A in the sample. Furthermore

$$A_0^2 = E_0\left[\left(\frac{x}{p_0} - \frac{1 - x}{1 - p_0}\right)^2\right] = [p_0(1 - p_0)]^{-1}.$$

95 per cent fiducial limits for p_0 are got by solving the following equation for p_0

$$\frac{m - np_0}{\sqrt{n}\sqrt{p_0(1 - p_0)}} = \pm 1.96.$$

It will be seen that situations, such as frequently occur in genetics, where p may be a function of some other parameter θ, say $p = u(\theta)$, can be handled by simply replacing p_0 by $u(\theta_0)$ and solving for θ_0.

EXAMPLE 3. Let the form of the distribution function be $\theta e^{-\theta x}$, where $0 \leq x < \infty$. For a sample of individuals,

$$L = n \log \theta - \theta \Sigma x_i$$

$$\left(\frac{\partial L}{\partial \theta}\right)_0 = \frac{n}{\theta_0} - \Sigma x_i$$

$$A_0^2 = E_0\left[\left(\frac{1}{\theta_0} - x\right)^2\right] = \frac{1}{\theta_0^2}.$$

The 95 per cent fiducial limits $\underline{\theta}$ and $\bar{\theta}$ are given by solving the equations

$$\frac{\dfrac{n}{\theta_0} - \Sigma x_i}{\sqrt{n}\,(1/\theta_0)} = \pm 1.96$$

for θ_0. We get

$$\underline{\theta} = \frac{1 - 1.96/\sqrt{n}}{\bar{x}}; \qquad \bar{\theta} = \frac{1 + 1.96/\sqrt{n}}{\bar{x}}$$

where $\bar{x}$ is the mean of the sample.

PRINCETON UNIVERSITY.

REFERENCES

[1] R. A. FISHER, "The Concepts of Inverse Probability and Fiducial Probability Referring to Unknown Parameters," *Proc. Royal Society of London*, Series A, vol. 139 (1933), pp. 343–348.

[2] R. A. FISHER, "The Fiducial Argument in Statistical Inference," *Annals of Eugenics*, vol. 6 (1935), pp. 391–398.

[3] J. NEYMAN, "On the Two Different Aspects of the Representative method: the Method of Stratified Sampling and the Method of Purposive Selection," *Royal Statistical Society*, vol. 97, 1934, pp. 558–625.

[4] J. NEYMAN, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," *Phil. Trans. Roy. Soc. London*, Series A, vol. 236 (1937), pp. 333–380.

[5] H. CRAMÉR, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 36, Cambridge University Press, 1937.

TRANSFORMATIONS OF THE PEARSON TYPE III DISTRIBUTION

By A. C. Olshen

I. Introductory

Transformations of the normal curve have been used as a basis for the representation of skew frequency distributions by Edgeworth, Kapteyn, Van Uven, Bernstein, and others. Various studies have been made of the distributions obtained by replacing each of a set of normally distributed variates by a logarithmic function of the variates. Among the earlier investigators along this line were Galton, and McAllister; later, works by Jorgensen, Fisher, Wicksell, Davies, and a more recent study by Pae-Tsi-Yuan, were added.

Rietz[1] restated and treated, in a general fashion, the question as to the properties of the distribution of powers of a set of variates which are known to be normally distributed. By a suitable choice for the origin of the normal curve, he obtained results which are applicable in answering questions which frequently arise in the applied field concerning the properties of families of interrelated distributions, one strain of which is known to be normally distributed. For example, in the family made up of the diameters, surface areas, volumes, etc. of some physical quantity, if it were known that one set, the surface areas for instance, were distributed normally, then from his results we have the properties of the distributions of any of the other sets.

Likewise it has seemed of interest to investigate, in a similar fashion, the properties of the transformed Type III Pearson distribution. We shall treat both the power and logarithmic transformations. For instance, if we knew that any one of the physical measurements, velocity, kinetic energy, momentum, or centrifugal force (all of which are functions of the velocity) were distributed according to a Type III curve, then we raise the question as to the properties of the distributions of any of the others. Similarly, if the intensity of certain light, I, were known to be distributed according to a Type III law, we will discuss the properties of the distribution of the brightness, B, of the light as seen by the eye, since the two are known to be related by the law $B = K \log I$. The same analysis applies to the relationship between L, the loudness of a sound, and E, the energy in the sound wave, since $L = K \log E$.

Two forms of the Type III distribution will be considered. In the first form, all the variates are taken positive; in the second form, the origin is at the mean and the variates are measured in units of standard deviation.

---

[1] H. L. Rietz, Frequency Distributions Obtained By Certain Transformations of Normally Distributed Variates, Annals of Math., Vol. 23, (1922) pp. 291–300.

In the last section, a transformation is developed which will transform the ordinates of a given probability function into the ordinates of the normal curve, $y = Ce^{\frac{-t^2}{2}}$, to within certain approximations. This transformation is applied to the Type III distribution and to the distribution obtained under power transformations of variates of the Type III distribution.

II. Power Transformations

a. *Type III curve with all variates positive.*

Given the Pearson Type III law,

$$y = y_0 x^{\gamma\bar{x}-1} e^{-\gamma x}, \qquad 0 \leq x < \infty, \tag{1}$$

where

$$\gamma = \frac{2\mu_2}{\mu_3} > 0, \qquad y_0 = \frac{(\gamma)^{\gamma\bar{x}}}{\Gamma(\gamma\bar{x})}, \qquad \bar{x} = \mu_1' > \frac{1}{\gamma}, \qquad x_{\text{mo.}} = \bar{x} - \frac{1}{\gamma}.\text{[2]} \tag{2}$$

The probability function (1) is a single-valued, real-valued, non-negative, continuous function of x with $\int_0^\infty y\,dx = 1$. The probability that a variate chosen at random will fall into the interval α_1 to α_2 is given by

$$P = \int_{\alpha_1}^{\alpha_2} y\,dx. \tag{3}$$

Let us make a transformation by replacing each variate x by x', where $x' = x^n$, and n is a real number on which restrictions will be placed as we proceed. When n is such that x' may have more than one value corresponding to an assigned value of x we shall consider only the principal value of x'. Then $dx' = nx^{n-1}\,dx$, except at $x = 0$ when $n < 1$, and $dx = \dfrac{dx'}{nx'^{\frac{n-1}{n}}}$ except at $x' = 0$ when $n > 1$, or $n < 0$.

The frequency function of the x' variates is given by

$$f(x') = \frac{y_0}{n} x'^{\frac{\gamma\bar{x}}{n}-1} e^{-\gamma x'^{\frac{1}{n}}}, \tag{4}$$

which does not represent a Type III curve when $n \neq 1$. The function (4) is discontinuous at $x' = 0$ if $\dfrac{\gamma\bar{x}}{n} < 1$. Likewise, corresponding to (3) we have

$$P = \frac{y_0}{n} \int_{\alpha_1^n}^{\alpha_2^n} x'^{\frac{\gamma\bar{x}}{n}-1} e^{-\gamma x'^{\frac{1}{n}}}\,dx'. \tag{5}$$

---

[2] The expression $x_{\text{mo.}}$ represents the mode, and $x_{\text{md.}}$ represents the median.

In order to study the maxima and minima points of (4) we take the derivative

$$\frac{df(x')}{dx'} = f(x')(nx')^{-1}\left\{-\gamma x'^{\frac{1}{n}} + \gamma \bar{x} - n\right\}. \tag{6}$$

The derivative changes signs at

$$x' = \left(\bar{x} - \frac{n}{\gamma}\right)^n. \tag{7}$$

Thus, variates in an interval, dx', at the mode of the new distribution (4) came by transformation, from an interval in the neighborhood of $x = \bar{x} - \frac{n}{\gamma}$, which is to the left of the mode of (1) when $n > 1$. The function (4) will be a monotone decreasing or a unimodal continuous distribution with mode given by (7) according as $\bar{x}$ is equal to or is greater than $\frac{n}{\gamma}$.

It will prove convenient to discuss the properties of (4) under three headings, according as $n > 1$, $0 < n < 1$, and $n < 0$, where n or its reciprocal is an integer.

Case I. $n > 1$.

When $\bar{x} < \frac{n}{\gamma}$, (4) is a monotone decreasing function, infinite at the origin and asymptotic to the x' axis; in this case the distribution of x' is similar to the distribution arising in the corresponding transformation of a set of normally distributed[8] variates, when $\bar{x} \leq 4(n-1)$, where $\bar{x}$ is the arithmetic mean of the x's of the normal curve. However, we are primarily interested in the case when $\bar{x} \geq \frac{n}{\gamma}$, under which condition a mode exists on the frequency curve $f(x')$ and is given by $x'_{\text{mo.}} = \left(\bar{x} - \frac{n}{\gamma}\right)^n$. Henceforth in discussing the comparative values of the measures of central tendency, it will be assumed that the condition $\bar{x} \geq \frac{n}{\gamma}$ is satisfied. We have,

$$\left(\bar{x} - \frac{n}{\gamma}\right)^n < \left(\bar{x} - \frac{1}{\gamma}\right)^n, \quad \text{where } x_{\text{mo.}} = \bar{x} - \frac{1}{\gamma}.$$

Thus, while variates at the modal value of x in the Type III distribution transform into $x' = \left(\bar{x} - \frac{1}{\gamma}\right)^n$, the mode of the new distribution is at $x'_{\text{mo.}} = \left(\bar{x} - \frac{n}{\gamma}\right)^n$ which, when $n > 1$, is to the left of the positions to which variates at the mode of the Type III distribution were transformed. Furthermore, as n increases, $x'_{\text{mo.}}$ approaches the origin.

---

[8] Cf. Rietz, loc. cit. p. 296.

The arithmetic mean of the x''s distributed in accord with the function (4) is given by

$$\bar{x}' = {}_{x'}\mu_1' = \frac{y_0}{n}\int_0^\infty x'^{\frac{\gamma\bar{x}}{n}} e^{-\gamma x'^{\frac{1}{n}}} dx' \tag{8}$$

$$= \frac{\Gamma(\gamma\bar{x}+n)}{\gamma^n\Gamma(\gamma\bar{x})}.$$

Similarly the sth moment about the origin is

$$_{x'}\mu_s' = \frac{\Gamma(\gamma\bar{x}+sn)}{\gamma^{sn}\Gamma(\gamma\bar{x})} = \frac{(\gamma\bar{x}+\overline{sn-1})^{(sn)}}{\gamma^{sn}}. \tag{9}$$

But,

$$\frac{\Gamma(\gamma\bar{x}+n)}{\gamma^n\Gamma(\gamma\bar{x})} = \left(\bar{x}+\frac{n-1}{\gamma}\right)\left(\bar{x}+\frac{n-2}{\gamma}\right)\cdots(\bar{x}),$$

which is greater than $(\bar{x})^n$, hence $\bar{x}' > \bar{x}^n$. Thus, while variates at the mean value of x in (1) transform into $(\bar{x})^n$, the mean of (4) is at $\dfrac{\Gamma(\gamma\bar{x}+n)}{\gamma^n\Gamma(\gamma\bar{x})}$ which is to the right of the positions to which variates at the mean of (1) were transformed. We have

$$\left(\bar{x}-\frac{n}{\gamma}\right)^n < (\bar{x})^n < \frac{\Gamma(\gamma\bar{x}+n)}{\gamma^n\Gamma(\gamma\bar{x})}, \tag{10}$$

hence

$$x'_{\text{mo.}} < \bar{x}'. \tag{11}$$

In 1895, Karl Pearson[4] showed that the median of the Type III curve was approximately two-thirds of the distance from the mode to the mean, and later Doodson[5] gave similar results. The analysis of (1) along this line is given in Section IV. However, since $x_{\text{mo.}} = \bar{x} - \dfrac{1}{\gamma}$, we may take $x_{\text{md.}} = \bar{x} - \dfrac{1}{c\gamma}$ where $c > 1$, (approximately equal to 3). Then $x'_{\text{md.}} = \left(\bar{x} - \dfrac{1}{c\gamma}\right)^n$. We have

$$\left(\bar{x}-\frac{n}{\gamma}\right)^n < \left(\bar{x}-\frac{1}{c\gamma}\right)^n < (\bar{x})^n, \tag{12}$$

hence from (10)

$$x'_{\text{mo.}} < x'_{\text{md.}} < \bar{x}'. \tag{13}$$

---

[4] Karl Pearson, Skew Variation in Homogeneous Material, Philosophical Transactions, Vol. 186A, part 1, (1895) pp. 343–414.

[5] Arthur T. Doodson, Relation of Mode, Median and Mean in Frequency Curves, Biometrika, Vol. 11, (1917) p. 425.

Considering the case when $n = 2$, we have with the aid of (9),

$$_{x'}\beta_1 = \frac{8(5\gamma^2\bar{x}^2 + 17\gamma\bar{x} + 15)^2}{(\gamma\bar{x})(\gamma\bar{x}+1)(2\gamma\bar{x}+3)^3},$$

and

$$_{x'}\beta_2 = \frac{3(4\gamma^4\bar{x}^4 + 72\gamma^3\bar{x}^3 + 337\gamma^2\bar{x}^2 + 620\gamma\bar{x} + 420)}{(\gamma\bar{x})(\gamma\bar{x}+1)(2\gamma\bar{x}+3)^2}.$$

From the moments of (1), one readily gets

$$_{x}\beta_1 = \frac{4}{\gamma\bar{x}},$$

and

$$_{x}\beta_2 = \frac{3(\gamma\bar{x}+2)}{\gamma\bar{x}}.$$

It can be shown easily that $_{x'}\beta_1 > {}_{x}\beta_1$ and $_{x'}\beta_2 > {}_{x}\beta_2$; hence the distribution of the squares of variates is more leptokurtic and more skew than the original distribution.

From (10) and (12) it is evident that the mode approaches neither the median nor the mean as n increases, subject to the condition $\gamma\bar{x} \geq n$. Each of the ratios of the mode to the median and to the mean approaches the limit 1 as $\bar{x}$ is increased indefinitely, the rapidity of approach to the limiting value depending on the size of n.

Taking the second derivative to find the points of inflection of the function (4), we have

$$\frac{d^2 f(x')}{dx'^2} = f(x')(nx')^{-2}\left\{\gamma^2 x'^{\frac{2}{n}} + \gamma x'^{\frac{1}{n}}(3n - 2\gamma\bar{x} - 1) + (\gamma^2\bar{x}^2 - 3n\gamma\bar{x} + 2n^2)\right\}.$$

When the points of inflection exist they are given by

$$x' = \frac{(2\gamma\bar{x} - 3n + 1) \pm \sqrt{n^2 - 6n + 1 + 4\gamma\bar{x}}}{2\gamma}. \tag{14}$$

Under the restriction that $\gamma\bar{x} \geq n$, the expression under the radical in (14) cannot vanish, and will always be positive.

Case II. $0 < n < 1$.

We now consider the distribution obtained by taking positive integral roots of a set of variates distributed in accord with (1). The mode of $f(x')$, as given by (7), will always exist since from (2), $\gamma\bar{x} > 1 > n$. We have

$$\begin{aligned} \left(\bar{x} - \frac{n}{\gamma}\right)^n &< \left(\bar{x} - \frac{1}{c\gamma}\right)^n \qquad \text{if } n > \frac{1}{c}, \\ \left(\bar{x} - \frac{n}{\gamma}\right)^n &> \left(\bar{x} - \frac{1}{c\gamma}\right)^n \qquad \text{if } n < \frac{1}{c}, \\ \left(\bar{x} - \frac{n}{\gamma}\right)^n &= \left(\bar{x} - \frac{1}{c\gamma}\right)^n \qquad \text{if } n = \frac{1}{c}. \end{aligned} \tag{15}$$

Hence it is evident that $x'_{mo.}$ is less than, greater than, or equal to $x'_{md.}$ according as n is greater than, less than, or equal to $1/c$. In power transformations of symmetrical distributions[6] the results differ in that the modal value is always greater or less than the median according as the value of n lies between 0 and 1 or outside of these bounds.

Here, $\left(\bar{x} - \frac{n}{\gamma}\right)^n > \left(\bar{x} - \frac{1}{\gamma}\right)^n$, hence in contrast to Case I, the mode of the new distribution is to the right of the position to which variates at the mode of the Type III distribution were transformed.

It has been shown[7] for every set of positive values that $\mu'_n > (\bar{x})^n$ when n lies outside of the interval 0 to 1, and that $\mu'_n < (\bar{x})^n$ when n lies between 0 and 1. We have then

$$\frac{\Gamma(\gamma\bar{x} + n)}{(\gamma)^n\Gamma(\gamma\bar{x})} = {}_{x'}\mu'_1 = {}_{x}\mu'_n < (\bar{x})^n. \tag{16}$$

The mean of the new distribution is to the left of the position to which the variates at the mean of the Type III distribution were transformed.

In Section IV, with the aid of certain approximating assumptions, it will be shown that $\bar{x}' > x'_{md.}$ when $x'_{md.} > x'_{mo.}$ and conversely, $\bar{x}' < x'_{md.}$ when $x'_{md.} < x'_{mo.}$

Case III. $n < 0$.

Let $n = -m$, where m is a positive integer. Then we have

$$f(x') = \frac{y_0}{m} x'^{-\left(\frac{\gamma\bar{x}}{m}+1\right)} e^{-\gamma x'^{-\frac{1}{m}}}, \qquad f(x') = 0 \quad \text{at} \quad x' = 0. \tag{17}$$

In place of (6) we have

$$\frac{df(x')}{dx'} = f(x')\left(mx'^{\frac{m+1}{m}}\right)^{-1}\left\{(\gamma\bar{x} + m)\,x'^{\frac{1}{m}} - \gamma\right\}, \tag{18}$$

and (7) becomes

$$x' = \left(\frac{1}{\bar{x} + \frac{m}{\gamma}}\right)^m. \tag{19}$$

But

$$\left(\frac{1}{\bar{x} + \frac{m}{\gamma}}\right)^m < \left(\frac{1}{\bar{x} - \frac{1}{\gamma}}\right)^m,$$

and

$$\bar{x}' = \frac{1}{\left(\bar{x} - \frac{m}{\gamma}\right)\left(\bar{x} - \frac{m-1}{\gamma}\right)\cdots\left(\bar{x} - \frac{1}{\gamma}\right)} > \left(\frac{1}{\bar{x}}\right)^m.$$

---

[6] H. L. Rietz, On Certain Properties of Frequency Distributions, Proc. National Academy of Science, Vol. 13, (1927) p. 820.

[7] J. L. W. V. Jensen, On Convex Functions and the Inequalities between the Means, Acta Mathematica, Vol. 30, pp. 175–193.

Hence, as in Case I, the mode of (17) is to the left of the position to which variates at the mode of (1) were transformed while the mean of (17) is to the right of the position to which the variates at the mean of (1) were transformed. Since

$$\left(\frac{1}{\bar{x}+\frac{m}{\gamma}}\right)^m < \left(\frac{1}{\bar{x}-\frac{1}{c\gamma}}\right)^m,$$

we have $x'_{\text{mo.}} < x'_{\text{md.}}$. Also

$$\frac{1}{\left(\bar{x}-\frac{m}{\gamma}\right)\left(\bar{x}-\frac{m-1}{\gamma}\right)\cdots\left(\bar{x}-\frac{1}{\gamma}\right)} > \left(\frac{1}{\bar{x}-\frac{1}{c\gamma}}\right)^m,$$

hence $\bar{x}' > x'_{\text{md.}}$. Therefore

$$x'_{\text{mo.}} < x'_{\text{md.}} < \bar{x}'. \tag{20}$$

As a special case, when $n = -1$, (17) reduces to

$$f(x') = y_0 x'^{-(\gamma\bar{x}+1)} e^{-\frac{\gamma}{x'}}, \tag{21}$$

which is a Pearson Type V distribution.

b. *Type III curve with mean zero and unit variance.*

Even though the form of the Type III distribution with which we have been dealing, wherein all the variates are positive, is more closely akin to actual distributions that may arise in applied problems, nevertheless it will be of interest to examine the properties of the transformed curve when the mean is taken as zero, with unit variance.

The second and third moments about the mean of the distribution (1) are ${}_x\mu_2 = \frac{\bar{x}}{\gamma}$, ${}_x\mu_3 = \frac{2\bar{x}}{\gamma^2}$. If we write α_3 for the third standard moment $\frac{\mu_3}{\mu_2^{3/2}}$, then

$$\gamma = \frac{4}{\bar{x}\alpha_3^2}. \tag{22}$$

By replacing the variable, x, in (1) by the expression

$$x = \bar{x}\left(1+\frac{\alpha_3 t}{2}\right), \tag{23}$$

we obtain the Type III distribution

$$\eta = \eta_0\left(1+\frac{\alpha_3 t}{2}\right)^{\frac{4}{\alpha_3^2}-1} e^{-\frac{2}{\alpha_3}t}, \tag{24}$$

where

$$\eta_0 = \frac{\left(\frac{4}{\alpha_3^2}\right)^{\frac{4}{\alpha_3^2}-\frac{1}{2}}}{e^{\frac{4}{\alpha_3^2}}\,\Gamma\left(\frac{4}{\alpha_3^2}\right)}, \qquad \alpha_3 < 2, \qquad -\frac{2}{\alpha_3} \le t < \infty, \qquad \bar{t} = 0, \qquad {}_t\mu_2 = 1.$$

Equation (22) lends itself to a simple interpretation of the restriction made in Section IIa, Case I, that $\gamma\bar{x} \geq n$ in order that the mode of (4) exist and be given by (7). The upper bound in the values of α_3 considered by Salvosa[8] in the computation of his tables was $\alpha_3 = 1.1$. Upon examination of the tables it is obvious that in most cases the skewness of the Type III distribution, as measured by α_3, will be less than 1.1. Hence in most cases we will have $3.22 \leq \gamma\bar{x} < \infty$. The effect of the limitations imposed by the condition $\gamma\bar{x} \geq n$ may be inferred to some extent from the following table.

TABLE I

The upper bound of α_3 for the existence of a mode in Case I, Section IIa

n	2	3	4	5	6	7	8	9	10	25	50	100
α_3	1.41	1.15	1.00	.89	.82	.76	.71	.67	.63	.40	.28	.20

When we make a transformation by replacing each variate t in (24) by t' where $t' = t^n$ $(n \neq 0)$ and n is an integer (positive or negative) or the reciprocal of an integer, then $dt' = nt^{n-1}\,dt$, except at $t = 0$ when $n < 1$, and $dt = \dfrac{dt'}{nt'^{\frac{n-1}{n}}}$, except at $t' = 0$ when $n > 1$, or $n < 0$. The function (24) becomes

$$f(t') = \frac{\eta_0}{n}\,\frac{\left(1 + \frac{\alpha_3}{2}t'^{\frac{1}{n}}\right)^{\frac{4}{\alpha_3^2}-1} e^{-\frac{2}{\alpha_3}t'^{\frac{1}{n}}}}{t'^{\frac{n-1}{n}}}. \tag{25}$$

The distribution function, $f(t')$, is infinite at $t' = 0$ when $n > 1$. In place of (5), we have

$$P = \frac{\eta_0}{n}\int_{a_1^n}^{a_2^n} \frac{\left(1 + \frac{\alpha_3}{2}t'^{\frac{1}{n}}\right)^{\frac{4}{\alpha_3^2}-1} e^{-\frac{2}{\alpha_3}t'^{\frac{1}{n}}}}{t'^{\frac{n-1}{n}}}\,dt'. \tag{26}$$

Here a_1 and a_2 are taken to be positive or zero when n is even. When n is odd a_1 and a_2 may be taken negative as (25) will give the frequency curve for negative values of t' that arise from setting $t' = t^n$ when t is negative. Examining for maxima and minima points, we have

$$\frac{df(t')}{dt'} = -f(t')\left\{nt'\left(1 + \frac{\alpha_3}{2}t'^{\frac{1}{n}}\right)\right\}^{-1}\left\{t'^{\frac{2}{n}} + \frac{n\alpha_3 t'^{\frac{1}{n}}}{2} + (n-1)\right\}. \tag{27}$$

[8] Luis R. Salvosa, Tables of Pearson's Type III Function, Annals of Mathematical Statistics, Vol. 1, (1930) pp. 191–198.

The derivative changes signs at

$$(28) \qquad t' = \frac{1}{2^n}\left(\frac{-n\alpha_3}{2} \pm \sqrt{\left(\frac{n\alpha_3}{2}\right)^2 - 4(n-1)}\right)^n$$

when $\left(\frac{n\alpha_3}{2}\right)^2 > 4(n-1)$, and at $t' = 0$ for certain values of n. When $n > 1$, $\left(\frac{n\alpha_3}{2}\right)^2 > 4(n-1)$ only for very large values of n since $\alpha_3 < 2$. When n is odd and positive, the derivative changes signs at $t' = 0$. If n is the reciprocal of an odd positive integer greater than one, there is a minimum at $t' = 0$, and the function $f(t')$ is zero at this point. Further properties of the frequency curves given by (25) will be discussed under the three cases treated in Section IIa.

Case I. $n > 1$.

When $\left(\frac{n\alpha_3}{2}\right)^2 > 4(n-1)$, it can be shown that (28) gives neither a maximum nor minimum point of $f(t')$ since $t'^{\frac{1}{n}}$ will always be less than t_l.[9] Similarly, when $\left(\frac{n\alpha_3}{2}\right)^2 < 4(n-1)$, there is neither a maximum nor minimum since (28) is imaginary. When n is odd, $f(t')$ is infinite at the origin and is a monotone increasing function of t' from the lower bound to the origin, and a monotone decreasing function of t' from the origin. When n is even, $f(t')$ is a monotone decreasing function of t', infinite at the origin. The forms of the distributions in this case are similar to those arising in power transformations of normally distributed variates[10] when $n > 1$ and $\bar{x}^2 \leq 4(n-1)$ and also to the forms arising in Section IIa, Case I, when $\bar{x} < \frac{n}{\gamma}$.

Even though we have a discontinuity at the origin, the total area under the curve is one, which is evident since we can integrate function (25) over the entire range of t' when n is odd and positive.

Case II. $0 < n < 1$.

This case includes the distribution obtained by taking positive integral roots of a set of variates. As in the study of the normal distribution,[11] we limit our considerations to the principal real values of the functions. When n is odd, there is a minimum at $t' = 0$ and a maximum given by each of the two signs before the radical in (28). Hence in this case, we have one minimum and two maxima.

With the values for n and α_3 in (24), $t_{md.} = -.164$ and $t_{mo.} = -.500$. The transformed distribution gives $t'_{md.} = -.547$ and two modes, the primary mode $t'_{mo.} = -.967$, and the secondary mode $t'_{mo.} = .903$. In contrast to the corresponding transformation of normally distributed variates, the primary mode is less than the median.

---

[9] The expression t_l represents the lower bound of t in distribution (24).

[10] H. L. Rietz, Cf. loc. cit. p. 296.

[11] Cf. Rietz, loc. cit. p. 297.

TABLE II

Comparison of the Type III Distribution and the Transformed Distribution when $\alpha_3 = .6,\ n = 3$

t	η	t'	$f(t')$
−3.00	.000001	−27.000000	0
−2.50	.001347	−15.625000	.000072
−2.00	.029467	− 8.000000	.002456
−1.75	.072787	− 5.359375	.007922
−1.50	.139285	− 3.375000	.020635
−1.25	.220462	− 1.953125	.047032
−1.00	.301350	− 1.000000	.100450
− .50	.405345	− .125000	.540460
− .25	.414211	− .015625	2.209125
− .10	.406131	− .001000	13.537700
− .05	.401485	− .000125	53.531333
− .02	.398272	− .000008	331.893333
0	.395962	0	∞
.02	.393522	.000008	327.935000
.05	.389628	.000125	51.950400
.10	.382549	.001000	12.751633
.15	.374795	.003375	5.552519
.25	.357533	.015625	1.906843
.50	.307293	.125000	.409724
.75	.252971	.421875	.149909
1.00	.200493	1.000000	.066831
1.50	.114233	3.375000	.016923
2.00	.058376	8.000000	.004865
2.50	.027285	15.625000	.001455
3.00	.011836	27.000000	.000438
3.50	.004820	42.875000	.000131
4.00	.001859	64.000000	.000039
5.00	.000242	125.000000	.000003
6.00	.000027	216.000000	0

Case III. $n < 0$.

Let $n = -m$, where m is a positive integer. Then (25) becomes

$$f(t') = \frac{y_0}{m} \frac{\left(1 + \frac{\alpha_3}{2} t'^{-\frac{1}{m}}\right)^{\frac{4}{\alpha_3^2}-1} e^{-\frac{2}{\alpha_3} t'^{-\frac{1}{m}}}}{t'^{\frac{m+1}{m}}}. \tag{29}$$

TABLE III

Comparison of the Type III Distribution and the Transformed Distribution When $\alpha_3 = 1, n = 1/3$

t	η	t'	$f(t')$
−2.00	0	−1.259921	0
−1.75	.025272	−1.205071	.11010
−1.50	.122826	−1.144714	.48208
−1.25	.251021	−1.077217	.87385
−1.00	.360894	−1.000000	1.08268
− .90	.393277	− .965489	1.00998
− .75	.427526	− .908560	1.05874
− .50	.448084	− .793701	.84683
− .27	.433958	− .646330	.54386
− .08	.405678	− .430887	.22506
0	.390734	0	0
.08	.374636	.430887	.20861
.27	.332027	.646330	.41723
.50	.280748	.793701	.53058
.64	.249865	.861774	.55069
.74	.228711	.904504	.56134
.90	.196904	.965489	.55084
1.00	.178470	1.000000	.53541
1.50	.104259	1.144714	.40985
2.00	.057252	1.259921	.27265
2.50	.029989	1.357209	.16572
3.00	.015133	1.442250	.09443
3.50	.007410	1.518295	.05125
4.00	.003539	1.587401	.02675
4.50	.001655	1.650964	.01353
5.00	.000761	1.709976	.00668
5.50	.000344	1.765174	.00322

Taking the derivative,

$$(30)\quad \frac{df(t')}{dt'} = -f(t')\left\{mt'^{\frac{m+2}{m}}\left(1 + \frac{\alpha_3}{2}t'^{-\frac{1}{m}}\right)\right\}^{-1}\left\{(m+1)t'^{\frac{2}{m}} + \frac{m}{2}\alpha_3 t'^{\frac{1}{m}} - 1\right\},$$

and in place of (28) we have

$$(31)\qquad t' = \left\{\frac{\dfrac{-m\alpha_3}{2} \pm \sqrt{\left(\dfrac{m\alpha_3}{2}\right)^2 + 4(m+1)}}{2(m+1)}\right\}^m.$$

The transformed distribution has little statistical significance for odd values of m, since $f(t')$ is a disjoined distribution. There are no values for $f(t')$ in the

interval, $\left(\frac{-\alpha_3}{2}\right)^m < t' < 0$, since $\frac{-2}{\alpha_3} \leq t < \infty$. The transformed distribution is thus composed of two sections, each with its own mode. The section for negative values of t', with range $-\infty < t' \leq \left(\frac{-\alpha_3}{2}\right)^m$, has a mode given by (31) with the negative sign before the radical. The section for positive values of t', with range $0 \leq t' < \infty$, has a mode given by (31) with the positive sign before the radical.

When m is an even integer, if we assign to $f(t')$ the value 0 when $t' = 0$, $f(t')$ becomes a continuous unimodal distribution in the interval $0 \leq t' < \infty$, with the mode given by (31), with the positive sign before the radical.

III. Logarithmic Transformations

As indicated in the introduction, numerous studies have been made of the distributions obtained by replacing normally distributed variates by exponential functions of the variates. If a variate x, with range $-\infty < x < \infty$, is distributed normally with mean zero and unit variance, then by replacing x by x', where $x' = c + e^{kx}$ the range of x' becomes, $c \leq x' < \infty$. Likewise if a variate x is distributed in accord with a Type III law, with range $0 \leq x < \infty$, then in making the above transformation, the range of x' becomes $(c + 1) \leq x' < \infty$. We shall now study the properties of the distribution of x' obtained by the above transformation applied to distribution (1). Because of the similarity of the properties of the transformed frequency distributions, we shall take $k = 1$ and $c = 0$.

Letting $x' = e^x$, (1) is transformed into

$$(32) \qquad f(x') = y_0 (\log x')^{\gamma\bar{x}-1} x'^{-(\gamma+1)}, \qquad 1 \leq x' < \infty.$$

Then,

$$(33) \qquad \frac{df(x')}{dx'} = f(x')\{x' \log x'\}^{-1}\{(\gamma\bar{x} - 1) - (\gamma + 1) \log x'\}.$$

The derivative changes signs at

$$(34) \qquad x' = e^{\frac{\gamma\bar{x}-1}{\gamma+1}}.$$

The arithmetic mean of the x''s distributed in accord with (32) is given by

$$\bar{x}' = y_0 \int_1^\infty (\log x')^{\gamma\bar{x}-1} x'^{-\gamma} \, dx'$$

$$= y_0 \int_0^\infty x^{\gamma\bar{x}-1} e^{-x(\gamma-1)} \, dx$$

$$(35) \qquad = \left(\frac{\gamma}{\gamma - 1}\right)^{\gamma\bar{x}} \quad \text{if} \quad \gamma > 1.$$

The integral is divergent when $0 < \gamma \leq 1$, hence for these cases we take $0 < k < \gamma$, then

$$\bar{x}' = y_0 \int_0^{\infty} x^{\gamma \bar{x}-1} e^{-x(\gamma-k)}\, dx$$

$$= \left(\frac{\gamma}{\gamma - k}\right)^{\gamma \bar{x}}. \tag{36}$$

Likewise in order that the first s moments about the origin be finite when $0 < \gamma \leq s$, we must have $0 < k < \frac{\gamma}{s}$.

The median of the distribution of x''s corresponds to

$$x = \log x' = \bar{x} - \frac{1}{c\gamma},$$

hence

$$x'_{\text{md.}} = e^{\bar{x} - \frac{1}{c\gamma}}. \tag{37}$$

1. **The relative positions of the averages.** We have $e^{\frac{\gamma \bar{x} - 1}{\gamma + 1}} < e^{\bar{x} - \frac{1}{c\gamma}}$ since $\frac{\gamma \bar{x} - 1}{\gamma + 1} < \bar{x} - \frac{1}{c\gamma}$. Hence

$$x'_{\text{mo.}} < x'_{\text{md.}}. \tag{38}$$

Also,

$$\gamma \bar{x} \log\left(\frac{\gamma}{\gamma - 1}\right) = \bar{x}\left[1 + \frac{1}{2\gamma} + \frac{1}{3\gamma^2} + \cdots\right] > \bar{x} - \frac{1}{c\gamma}.$$

Therefore

$$e^{\bar{x} - \frac{1}{c\gamma}} < \left(\frac{\gamma}{\gamma - 1}\right)^{\gamma \bar{x}}$$

and hence

$$x'_{\text{md.}} < \bar{x}'. \tag{39}$$

From (38) and (39), we have

$$x'_{\text{mo.}} < x'_{\text{md.}} < \bar{x}'. \tag{40}$$

We shall now investigate the locations of the various averages as related to the upper and lower points of inflection whose abscissas will be denoted by I'_u and I'_l respectively. Taking the second derivative,

$$\frac{d^2 f(x')}{dx'^2} = f(x')\, \{x' \log x'\}^{-2} \{(\gamma \bar{x} - 1)(\gamma \bar{x} - 2) - (2\gamma^2 \bar{x} + 3\gamma \bar{x} - 2\gamma - 3) \log x'$$

$$+ (\gamma + 1)(\gamma + 2) \ldots$$

The points of inflection are given by

(41) $$x' = e^{\varphi(\gamma, \bar{x})},$$

where

$$\varphi(\gamma, \bar{x}) = \frac{(\gamma\bar{x}-1)(2\gamma+3) \pm \sqrt{(\gamma\bar{x}-1)\{(\gamma\bar{x}-1)+4(\gamma+1)(\gamma+2)\}}}{2(\gamma+1)(\gamma+2)}.$$

(a). To show $x'_{\text{mo.}} > I'_l$.

We have,

$$\frac{\gamma\bar{x}-1}{\gamma+1} > \frac{(\gamma\bar{x}-1)(2\gamma+3)-(\gamma\bar{x}-1)(p+1)}{2(\gamma+1)(\gamma+2)},$$

where

$$p+1 = \sqrt{1+\frac{4(\gamma+1)(\gamma+2)}{\gamma\bar{x}-1}}, \quad \text{since } 2\gamma+4 > 2\gamma+2-p.$$

Therefore

$$e^{\frac{\gamma\bar{x}-1}{\gamma+1}} > e^{\varphi_l(\gamma,\bar{x})}.$$

(b). To show $x'_{\text{mo.}} < I'_u$.

We have,

$$\frac{\gamma\bar{x}-1}{\gamma+1} < \frac{(\gamma\bar{x}-1)(2\gamma+3)+(\gamma\bar{x}-1)(p+1)}{2(\gamma+1)(\gamma+2)},$$

since $2\gamma+4 < 2\gamma+4+p$.

Therefore,

$$e^{\frac{\gamma\bar{x}-1}{\gamma+1}} < e^{\varphi_u(\gamma,\bar{x})}.$$

From (a) and (b), we have

(42) $$I'_l < x'_{\text{mo.}} < I'_u.$$

(c). To show $x'_{\text{md.}} > I'_l$.

We have,

$$x'_{\text{md.}} > x'_{\text{mo.}} \text{ and } x'_{\text{mo.}} > I'_l.$$

Consequently

$$x'_{\text{md.}} > I'_l.$$

(d). To show the conditions under which $x'_{\text{md.}}$ is less than or greater than I'_u.

Upon simplifying the inequality we find that $e^{\varphi_u(\gamma,\bar{x})}$ will be greater or less than $e^{\bar{x}-\frac{1}{c\gamma}}$ according as the expression

(43) $$-2\bar{x}^2+\bar{x}\left\{\frac{1}{c}\left(3+\frac{4}{\gamma}\right)+(\gamma-3)\right\}+\frac{1}{c}\left(2+\frac{3}{\gamma}\right)-\frac{1}{c^2\gamma}\left(3+\frac{2}{\gamma}\right)-2-\frac{1}{c^2}$$

is positive or negative. But (43) will be negative for all values of $\bar{x}$ if its discriminant is negative or zero. Upon further examination it can be seen that the discriminant will be negative or zero according as

$$(44) \qquad \gamma^2 - 6\gamma\left(1 - \frac{1}{c}\right) + \frac{1}{c}\left(6 + \frac{1}{c}\right) - 7 \leq 0.$$

The quadratic equation in γ, given by (44), factors into

$$\left(\gamma - \frac{A - B}{2}\right)\left(\gamma - \frac{A + B}{2}\right),$$

where

$$A = 6\left(1 - \frac{1}{c}\right), \qquad B = \sqrt{36\left(1 - \frac{1}{c}\right)^2 - \frac{4}{c}\left(6 + \frac{1}{c}\right) + 28}, \qquad B > A.$$

Hence in order that $x'_{\mathrm{md.}}$ be greater than I'_u for all values of $\bar{x}$, we must have

$$\gamma \leq \frac{1}{2}\left\{6\left(1 - \frac{1}{c}\right) + \sqrt{36\left(1 - \frac{1}{c}\right)^2 - \frac{4}{c}\left(6 + \frac{1}{c}\right) + 28}\right\}.$$

When c lies in the neighborhood of 3, γ must be less than 5. Proceeding further, we can divide (43) by negative 2, reverse the inequalities, and factor the expression into

$$\left(\bar{x} - \frac{A' - B'}{4}\right)\left(\bar{x} - \frac{A' + B'}{4}\right),$$

where

$$A' = \left\{\frac{1}{c}\left(3 + \frac{4}{\gamma}\right) + (\gamma - 3)\right\}$$

$$B' = \sqrt{A'^2 + 8\left\{\frac{1}{c}\left(2 + \frac{3}{\gamma}\right) - \frac{1}{c^2\gamma}\left(3 + \frac{2}{\gamma}\right) - 2 - \frac{1}{c^2}\right\}}$$

$$B' < A'.$$

Then

$$(45) \qquad x'_{\mathrm{md.}} > I'_u \quad \text{if} \quad \bar{x} < \frac{A' - B'}{4} \quad \text{or} \quad \bar{x} > \frac{A' + B'}{4},$$

and

$$(46) \qquad x'_{\mathrm{md.}} < I'_u \quad \text{if} \quad \frac{A' - B'}{4} < \bar{x} < \frac{A' + B'}{4}.$$

(e). To show $I'_1 < \bar{x}'$.

We have, $I_l' < \bar{x}'$, since $\bar{x}' > x'_{\text{md.}} > x'_{\text{mo.}}$ and $x'_{\text{mo.}} > I_l'$. Also, $I_u' < \bar{x}'$ for those values of γ and $\bar{x}$ for which $I_u' < x'_{\text{md.}}$. It remains to be shown that $\bar{x}'$ is always greater than I_u' for all values of γ and $\bar{x}$. To show that

$$e^{\varphi_u(\gamma,\bar{x})} < \left(\frac{\gamma}{\gamma-1}\right)^{\gamma\bar{x}},$$

it will be sufficient to show that $\varphi_u(\gamma, \bar{x}) < \bar{x}$, since

$$\gamma\bar{x}\log\left(\frac{\gamma}{\gamma-1}\right) = \bar{x}\left(1 + \frac{1}{2\gamma} + \frac{1}{3\gamma^2} + \cdots\right) > \bar{x}.$$

The inequality is satisfied if

$$(\gamma\bar{x}-1)\{(\gamma\bar{x}-1)+4(\gamma+1)(\gamma+2)\} < \{2\bar{x}(\gamma+1)(\gamma+2)-(\gamma\bar{x}-1)(2\gamma+3)\}^2.$$

This expression, however, reduces to the condition that we must have

$$-2-\gamma\bar{x}-6\bar{x}-6\gamma\bar{x}^2-\gamma^2\bar{x}-8\bar{x}^2 < 0,$$

which is always true. Hence

$$e^{\varphi_u(\gamma,\bar{x})} < \left(\frac{\gamma}{\gamma-1}\right)^{\gamma\bar{x}},$$

and we have

$$I_u' < \bar{x}'. \tag{47}$$

2. Contact at the ends of the range. We shall now investigate whether the frequency function, $f(x')$, has high contact with the x' axis. The function, $f(x')$, vanishes at both limits, and thus it will be sufficient to test the derivatives. The nth derivative can be expressed in the form of $f(x')(x' \log x')^{-n}$ multiplied by an nth degree polynomial in $\log x'$. It can easily be shown that each derivative will vanish at the upper limit, while the nth derivative will vanish at the lower limit provided $\gamma\bar{x} > n$. Therefore, $f(x')$ does not have high contact at the lower end of the range.

3. Moments. The sth moment about the origin is given by

$$
\begin{aligned}
{}_{x'}\mu_s' &= y_0 \int_1^\infty (\log x')^{\gamma\bar{x}-1} x'^{s-\gamma-1}\,dx' \\
&= y_0 \int_0^\infty x^{\gamma\bar{x}-1} e^{-x(\gamma-s)}\,dx \\
&= \left(\frac{\gamma}{\gamma-s}\right)^{\gamma\bar{x}}, \quad \text{if} \quad \gamma > s. \qquad (48)
\end{aligned}
$$

If $\gamma \leq s$, then taking k such that $\gamma > sk > 0$, we get in place of (48),

$${}_{x'}\mu_s' = \left(\frac{\gamma}{\gamma-sk}\right)^{\gamma\bar{x}}. \tag{49}$$

We easily obtain the recurring relationship

$$(50)\qquad {}_{x'}\mu_s' = \left(\frac{\gamma - s - 1}{\gamma - s}\right)^{\gamma z} {}_{x'}\mu_{s-1}'.$$

The sth moment of x' about the mean is

$$\begin{aligned} {}_{x'}\mu_s &= y_0 \int_1^\infty \left\{x' - \left(\frac{\gamma}{\gamma - 1}\right)^{\gamma z}\right\}^s (\log x')^{\gamma z - 1} x'^{-(\gamma+1)}\, dx' \\ &= y_0 \int_0^\infty \left\{e^x - \left(\frac{\gamma}{\gamma-1}\right)^{\gamma z}\right\}^s x^{\gamma z - 1} e^{-\gamma x}\, dx \end{aligned}$$

$$(51)\qquad = \left(\frac{\gamma}{\gamma - s}\right)^{\gamma z} \sum_{j=0}^{s} (-1)^j \binom{s}{j} \left(\frac{\gamma}{\gamma - 1}\right)^{j\gamma z} \left(\frac{\gamma - s}{\gamma - s - j}\right)^{\gamma z}.$$

If we do not take the value of k to be 1, then

$$(52)\qquad {}_{x'}\mu_s = \left(\frac{\gamma}{\gamma - sk}\right)^{\gamma z} \sum_{j=0}^{s} (-1)^j \binom{s}{j} \left(\frac{\gamma}{\gamma - k}\right)^{j\gamma z} \left(\frac{\gamma - sk}{\gamma - s - jk}\right)^{\gamma z}.$$

IV. Transformation into a Normal Distribution

We shall now consider a unimodal probability function $y = f(x)$ with range, $a \leq x \leq b$, and shall seek to express x as such a function of t as will transform $y = f(x)$, into $y = Ce^{-\frac{t^2}{2}}$. For simplicity, we assume that $y = f(x)$ has its modal value at $x = 0$, and thus each of the curves $y = f(x)$ and $y = Ce^{-\frac{t^2}{2}}$ has its one maximum value at the origin.

In $y = f(x)$ let $\log y = V$, then equating densities,[12] we have,

$$V - \log C + \frac{t^2}{2} = 0.$$

Then

$$\frac{dV}{dt} + t = 0,$$

$$(53)\qquad \frac{d^2V}{dt^2} + 1 = 0,$$

$$\frac{d^nV}{dt^n} = 0, \qquad n \geq 3.$$

---

[12] If $f(x)$ is a probability function or density of a distribution, then $f(x)\,dx$ is, to within infinitesimals of higher order, the probability that a value taken at random will fall into the interval dx at x.

Under the assumption that x is a function which can be expanded in a Maclaurin series in powers of t, we shall use equations (53) to determine the values of $\left.\frac{d^n x}{dt^n}\right]_{t=0}$ in the series

(54) $$x = A_0 + A_1 t + A_2 \frac{t^2}{2!} + \cdots$$

Let v_n represent $\left.\frac{d^n V}{dx^n}\right]_{t=0}$, $n = 0, 1, 2 \cdots$. Then $v_0 = \log C$, and $v_1 = 0$, since $\frac{dV}{dx}$ and $\frac{dy}{dt}$ vanish when $x = 0$, that is when $t = 0$, since A_0 is taken to be zero. Taking the second derivative,

$$\frac{d^2 V}{dt^2} = \frac{d^2 V}{dx^2}\left(\frac{dx}{dt}\right)^2 + \frac{d^2 x}{dt^2}\left(\frac{dV}{dx}\right) = -1,$$

and

$$\left.\frac{d^2 V}{dt^2}\right]_{t=0} = v_2 A_1^2 = -1.$$

Therefore we have

(55) $$A_1 = (-v_2)^{-\frac{1}{2}}, \text{ when } v_2 < 0.$$

Also,

$$\left.\frac{d^3 V}{dt^3}\right]_{t=0} = v_3 A_1^3 + 3v_2 A_1 A_2 = 0,$$

and we have

(56) $$A_2 = \frac{v_3}{3v_2^2}.$$

Similarly,

$$A_3 = \frac{5v_3^2 - 3v_4 v_2}{12v_2^4}(-v_2)^{\frac{1}{2}},$$

(57) $$A_4 = \frac{-(40v_3^3 - 45v_2 v_3 v_4 + 9v_2^2 v_5)}{45v_2^5},$$

$$A_5 = \frac{-\{385v_3^4 - 630v_2 v_3^2 v_4 + 21v_2^2(8v_3 v_5 + 5v_4^2) - 24v_2^3 v_6\}(-v_2)^{\frac{1}{2}}}{144v_2^7},$$

$$\vdots \qquad\qquad \vdots$$

Though the procedure is straightforward, the work becomes somewhat involved in determing A_n as n gets larger. For this reason, we proceed in the

following manner. By the use of Burmann's[15] theorem we can write (54) in the form

$$(58)\qquad x = \left\{\frac{x}{t}\right\}_{x=0} t + \left\{\frac{d}{dx}\left(\frac{x}{t}\right)^2\right\}_{x=0}\frac{t^2}{2!} + \left\{\frac{d^2}{dx^2}\left(\frac{x}{t}\right)^3\right\}_{x=0}\frac{t^3}{3!} + \cdots$$

But

$$\frac{x}{t} = \frac{x}{\sqrt{2}}(\log C - V)^{-\frac{1}{2}},$$

where V is a function of x. We have,

$$\log C = V\Big]_{x=0}, \qquad \frac{dV}{dx} = 0 \quad \text{at} \quad x = 0,$$

and we may write

$$V = \log C + \frac{d^2V}{dx^2}\Big]_{x=0}\frac{x^2}{2!} + \frac{d^3V}{dx^3}\Big]_{x=0}\frac{x^3}{3!} + \cdots$$

Hence,

$$\frac{x}{t} = \frac{x}{\sqrt{2}}(a_2x^2 + a_3x^3 + a_4x^4 + \cdots)^{-\frac{1}{2}}$$

$$= \frac{1}{\sqrt{2}}(a_2 + a_3x + a_4x^2 + \cdots)^{-\frac{1}{2}}, \quad \text{where} \quad a_n = -\frac{v_n}{n!}.$$

We can now write,

$$(59)\qquad \left(\frac{x}{t}\right)^n = \frac{1}{2^{\frac{n}{2}}}\left\{\sum_{s=0}^{\infty} a_{s+2}x^s\right\}^{-\frac{n}{2}}.$$

But

$$A_n = \frac{d^{n-1}}{dx^{n-1}}\left(\frac{x}{t}\right)^n\Big]_{x=0} = (n-1)! \text{ multiplied by the coefficient of } x^{n-1} \text{ in (59).}$$

Hence,

$$(60)\qquad A_n = (n-1)!\,(2a_2)^{-\frac{n}{2}}\sum \frac{\left(-\frac{n}{2}\right)\left(-\frac{n}{2}-1\right)\left(-\frac{n}{2}-2\right)\cdots\left(-\frac{n}{2}-p+1\right)}{\lambda_1!\lambda_2!\lambda_3!\lambda_n!} \left(\frac{a_3}{a_2}\right)^{\lambda_1}\left(\frac{a_4}{a_2}\right)^{\lambda_2}\left(\frac{a_5}{a_2}\right)^{\lambda_3}\cdots\left(\frac{a_{n+2}}{a_2}\right)^{\lambda_n}$$

where the summation is over all values of λ_s such that

$$\sum_{s=1}^{n} s\lambda_s = 0, \quad \text{and} \quad p = \sum_{s=1}^{n}\lambda_s.$$

[15] A. De Morgan, Differential and Integral Calculus, (1842) page 305.

This expression can be written in the form,

$$A_n = (n-1)!\,(2a_2)^{-\frac{n}{2}} \sum_{s=0}^{n-2} (-1)^{s+1} \frac{n}{2} \cdot \frac{n+2}{4} \cdot \frac{n+4}{6} \cdots \frac{n+2s}{2(s+1)} \frac{D^{n-s-2} a_3^{s+1}}{a_2^{s+1}}, \qquad n > 1, \tag{61}$$

where D is the derivative operator of Arbogast.[14]

If we take expression (1) as our function of x, we obtain,

$$v_n = (-1)^{n-1} \frac{(n-1)!\,\gamma^n}{(\gamma\bar{x}-1)^{n-1}},$$

which gives

$$x = \frac{(\gamma\bar{x}-1)}{\gamma} + \frac{(\gamma\bar{x}-1)^{\frac{1}{2}}}{\gamma} t + \frac{2}{3\gamma}\frac{t^2}{2!} + \frac{1}{6\gamma(\gamma\bar{x}-1)^{\frac{1}{2}}}\frac{t^3}{3!} + \frac{4}{45\gamma(\gamma\bar{x}-1)}\frac{t^4}{4!} + \frac{1}{36\gamma(\gamma\bar{x}-1)^{\frac{3}{2}}}\frac{t^5}{5!} + \cdots \tag{62}$$

where $A_n = (n-1)!\,\dfrac{(\gamma\bar{x}-1)^{\frac{n}{2}}}{2^{\frac{n}{2}}\gamma^n}$ multiplied by the coefficient of

$$\left[x - \left(\bar{x} - \frac{1}{\gamma}\right)\right]^{n-1} \quad \text{in} \quad \left\{\sum_{s=0}^{\infty} \frac{(-1)^s \gamma^s [x - (\bar{x} - 1/\gamma)]^s}{(s+2)(\gamma\bar{x}-1)^s}\right\}^{-\frac{n}{2}}$$

This series is known to diverge for large values of t. However, the series is defined for those values of t that correspond to x for the interval $0 < x < 2\left(\bar{x} - \dfrac{1}{\gamma}\right)$. With the aid of (22) and Salvosa's[15] tables we give in Table IV the percentage of the total population which is included in this interval.

TABLE IV

The percentage of the population, characterized by (1), which is included in the interval $0 < x < 2(\bar{x} - 1/\gamma)$, for different degrees of skewness

α_3	1.1	1.0	.9	.8	.7	.6	.5	.4	.3	.2	.1
Percent of Population	79.386	84.880	89.781	93.908	97.021	98.950	99.805	99.990	100.000	100.000	100.000

Thus, in dealing with samples as large as 10,000, with moderate degrees of skewness, the probability of getting a value that falls beyond this interval

[14] Augustus De Morgan, On Arbogast's Formulae of Expansion, Cambridge and Dublin Mathematical Journal, Vol. 1, (1848) pp. 238-255.

[15] Cf. Salvosa, loc. cit. page 2.

becomes negligible. Hence it may be expected that with the use of a comparatively few terms of series (62) we may transform the ordinates of a moderately skew Type III function to within close approximations of the ordinates of the normal function.

Baker[16] considered the transformations of a non-normal frequency distribution represented by $f(t)dt$, where the origin is taken at some central point and the scale is the standard deviation of the distribution. By equating probabilities he found a function φ, such that by setting $t = \varphi(u)$, he obtained

$$f[\varphi(u)].\quad \varphi'(u)\,du = e^{-\frac{1}{2}u^2}\,du.$$

It seemed of interest to compare the results obtained by applying transformation (62), which is found by equating densities, to the illustration treated by Baker,[17] where the transformation giving equality of probabilities was used.

The example treated was

$$(63)\qquad f(t) = .9929\left(1 + \frac{t}{10}\right)^{99} e^{-10t}.$$

This is a Type III distribution of form (24), with $\alpha_3 = .2$. From (22), $\gamma\bar{x} = \frac{4}{\alpha_3^2} = 100$, and from (62) we obtain the series

$$(64)\quad x = \frac{99}{\gamma}(1 + .1006038u + .0033670u^2 + .0000282u^3 - .0000004u^4 + \cdots).$$

We shall utilize only the first four terms and rewrite (64) in the form

$$\gamma x = 99(1 + .1006038u + .0033670u^2 + .0000282u^3).$$

However, from (23),

$$x = \bar{x}\left(1 + \frac{\alpha_3 t}{2}\right),$$

which gives

$$\gamma x = \gamma\bar{x}\left(1 + \frac{\alpha_3 t}{2}\right) = 100(1 + .1t).$$

Therefore,

$$t = .1(\gamma x - 100).$$

---

[16] G. A. Baker, Transformation of Non-normal Frequency Distributions into Normal Distributions, Annals of Mathematical Statistics, Vol. 5, (1934) pp. 113-123.

[17] Baker, loc. cit. page 117.

With the aid of Salvosa's[18] tables, we obtain the following results.

TABLE V

Comparison of the ordinates of the normal function, the function with skewness .2, and the skewed function transformed by

$$t = 9.9\,(-.0101010 + .1005038\,u + .0033670\,u^2 + .0000282\,u^3)$$

u	Normal Curve	Function with skewness .2	Transformed skew curve	Transformed skew curve *minus normal*
−2.0	.053991	.049243	.054226	.000235
−1.8	.078950	.076810	.079291	.000341
−1.6	.110921	.112956	.111393	.000472
−1.4	.149727	.157043	.150359	.000632
−1.2	.194186	.206951	.195003	.000817
−1.0	.241971	.259120	.242986	.001015
− .8	.289692	.308958	.290905	.001213
− .6	.333225	.351638	.334618	.001393
− .4	.368270	.382453	.369811	.001541
− .2	.391043	.398583	.392678	.001635
0	.398942	.398610	.400615	.001673
.2	.391043	.383157	.392682	.001639
.4	.368270	.354545	.369811	.001541
.6	.333225	.316273	.334621	.001396
.8	.289692	.272360	.290905	.001213
1.0	.241971	.226714	.242984	.001013
1.2	.194186	.182641	.194999	.000823
1.4	.149727	.142563	.150353	.000626
1.6	.110921	.107939	.111383	.000462
1.8	.078950	.079354	.079277	.000327
2.0	.053991	.056702	.054214	.000223

The ordinates of the transformed distribution are more symmetrical and approximate the ordinates of the normal curve more closely than the values obtained by Baker even though we have used only four terms in the transforming series.

Returning to the general case, we may write

$$\int_a^b y\,dx = \int_{-\infty}^{\infty} Ce^{-\frac{t^2}{2}}\frac{dx}{dt}\cdot dt$$

(65)

$$= C\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\left(A_1 + A_2 t + A_3\frac{t^2}{2!} + A_4\frac{t^3}{3!} + \cdots\right)dt,$$

[18] Salvosa, loc. cit. pp. 64 et seq.

provided the series converges for all values of t. Under the assumption that the integrand satisfies conditions for the term by term integration of the series, we get

$$(66)\quad 1 = \int_a^b y\,dx = C\sqrt{2\pi}\left\{A_1 + \frac{A_3}{2!} + \frac{3A_5}{4!} + \cdots + \frac{A_{2n+1}}{2^n n!} + \cdots\right\}.$$

The area to the right of the modal ordinate is

$$\int_{x_{mo.}}^b y\,dx = C\int_0^\infty e^{-\frac{t^2}{2}}\left(A_1 + A_2 t + \frac{A_3 t^2}{2!} + \cdots\right)dt$$

$$= \tfrac{1}{2} + C\int_0^\infty e^{-\frac{t^2}{2}}\left(A_2 t + A_4\frac{t^3}{3!} + \cdots\frac{A_{2n}t^{2n-1}}{(2n-1)!} + \cdots\right)dt$$

$$(67)\quad = \tfrac{1}{2} + C\left(A_2 + \frac{A_4}{3} + \cdots\right).$$

Hence the area from the mode to the median is

$$(68)\quad C\left(A_2 + \frac{A_4}{3} + \cdots\right).$$

Let us consider distribution (1) again. The coefficients in series (62) are functions of the skewness, and become smaller with smaller degrees of skewness. Indications are that with moderate skewness, the series converges sufficiently to be used for certain formal purposes. If we assume this and proceed in a formal manner we obtain some interesting results that are consistent with approximations that have been obtained elsewhere.

Thus, it is interesting to note that using the coefficients of series (62) in equation (66), we obtain

$$\Gamma(\gamma\bar{x}) = \sqrt{2\pi(\gamma\bar{x}-1)}\,(\gamma\bar{x}-1)^{\gamma\bar{x}-1}e^{-(\gamma\bar{x}-1)}$$
$$\left\{1 + \frac{1}{12(\gamma\bar{x}-1)} + \frac{1}{288(\gamma\bar{x}-1)^2} + \cdots\right\},$$

which is Stirling's asymptotic form for $\Gamma(\gamma\bar{x})$.[19]

From (68), the area from the mode to the median in the Type III distribution characterized by (1) is approximately

$$(69)\quad C\left(\frac{2}{3\gamma} - \frac{4}{135\gamma(\gamma\bar{x}-1)} + \cdots\right),$$

where

$$C = \frac{\gamma(\gamma\bar{x}-1)^{\gamma\bar{x}-1}e^{-(\gamma\bar{x}-1)}}{\Gamma(\gamma\bar{x})}.$$

---

[19] E. Czuber, Wahrscheinlichkeitsrechnung, Volume 1, (1908) pp. 23–24.

Since $(\gamma\bar{x} - 1)$ is large when the skewness is moderate, and since the terms of (69) are rapidly decreasing, the area from the mode to the median is approximately equal to $\frac{2C}{3\gamma}$. But $\frac{1}{\gamma}$ is the distance from the mode to the mean and C is the ordinate at the mode, hence the area from the mode to the median is approximately equal to the ordinate at the mode multiplied by 2/3 of the distance between the mode and mean. Therefore with moderate skewness the median is approximately 2/3 of the distance between the mode and mean, which conforms to the approximate result first obtained by Karl Pearson[20] for the Type III distribution. We may, for all cases resulting in (68), take A_2 as being approximately equal to the distance from the mode to the median. This becomes somewhat more apparent by finding the arithmetic mean of distribution y. Thus,

$$\mu_1' = \frac{\int_a^b xy\,dx}{\int_a^b y\,dx} = \frac{C\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\left(A_0 + A_1 t + \frac{A_2 t^2}{2!} + A_3\frac{t^3}{3!} + \cdots\right)\left(A_1 + A_2 t + A_3\frac{t^2}{2!} + A_4\frac{t^3}{3!} + \cdots\right)dt}{C\sqrt{2\pi}\left(A_1 + \frac{A_3}{2!} + \frac{3A_5}{4!} + \cdots\right)}$$

$$= \frac{A_0A_1 + \left(\frac{3A_1A_2}{2!} + \frac{A_0A_3}{2!}\right) + 3\left(\frac{A_0A_5}{4!} + \frac{5A_1A_4}{4!} + \frac{5A_2A_3}{2!\,3!}\right) + \cdots}{A_1 + \frac{A_3}{2!} + \frac{3A_5}{4!} + \cdots}$$

$$= A_0 + \frac{3A_2}{2!} + \frac{15A_4}{4!} + \cdots. \tag{70}$$

Remembering that A_0 is the abscissa of the mode, it becomes apparent that the mean is, in general, approximately equal to the mode plus 3/2 of the distance from the mode to the median.

Though series (62) is known not to converge for large values of t, it is interesting to note that if we use distribution (1) for y, we have from (70)

$$\mu_1' = \left(\bar{x} - \frac{1}{\gamma}\right) + \frac{3}{2}\left(\frac{2}{3\gamma}\right) - \frac{1}{3\,\gamma(\gamma\bar{x} - 1)\,3!} + \cdots, \tag{71}$$

the first two terms of which give $\bar{x}$, which is μ_1', and hence if (71) were an exact formula, the sum of the terms beyond the second would be zero.

For example (63), it can be seen from the following that A_2 furnishes a close approximation to the distance from the mode to the median. Here, $t = .1\,(\gamma\bar{x} - 100)$; putting $x = \bar{x} - \frac{1}{\gamma}$ we have $t_{mo.} = .1(\gamma\bar{x} - 101) = -.1$. Putting

[20] Karl Pearson, loc. cit.

$x = \left(\bar{x} - \frac{1}{\gamma}\right) + \frac{2}{3\gamma}$, where $A_2 = \frac{2}{3\gamma}$, we have as our approximation to the median

$$t_{md.} = .1\left(\gamma\bar{x} - \frac{301}{3}\right)$$
$$= -.03333.$$

Interpolating in the Salvosa tables, we find for $\alpha_3 = .2$, $t_{md.} = -.03331$ approximately. Hence it is seen that the interpolated values checks very closely with that obtained by using the A_2 criterion.

We shall now consider briefly the transforming series, when for y, we take distribution (4). Then, corresponding to (54), we obtain the series

$$(72) \qquad x' = \frac{(\gamma\bar{x} - n)^n}{\gamma^n} + \frac{n(\gamma\bar{x} - n)^{n-1}}{\gamma^n}t + \frac{n(3n-1)(\gamma\bar{x} - n)^{n-1}}{3\gamma^n}\frac{t^2}{2!}$$
$$+ \frac{n(6n^2 - 6n + 1)(\gamma\bar{x} - n)^{n-1}}{6\gamma^n}\frac{t^3}{3!}$$
$$+ \frac{n(45n^3 - 90n^2 + 45n - 4)(\gamma\bar{x} - n)^{n-2}}{45\gamma^n}\frac{t^4}{4!} + \cdots.$$

When $n = 1$, (72) reduces to the series given by (62). Suppose we are primarily interested in the cases for which $0 < n < 1$. For these cases the coefficients of (72) decrease more rapidly than do those of series (62). Under the same assumptions as to convergence which were made in working with the latter series we have, from (68) and (72), the area from the mode to the median given approximately by

$$(73) \quad C\left\{\frac{n(3n-1)(\gamma\bar{x} - n)^{n-1}}{3\gamma^n} + \frac{n(45n^3 - 90n^2 + 45n - 4)(\gamma\bar{x} - n)^{n-2}}{135\gamma^n} + \cdots\right\}.$$

When $0 < n < 1$ we always have $\gamma\bar{x} > n$; then $A_2 > 0$ if $n > 1/3$, and $A_2 < 0$ if $n < 1/3$. Therefore, if A_2 is taken to be approximately equal to the distance from the mode to the median, we have $x'_{mo.} > x'_{md.}$ if $n < 1/3$, and $x'_{mo.} < x'_{md.}$ if $n > 1/3$, since A_2 is positive or negative according as n is greater or less than 1/3. Combining these results with (70), we have $\bar{x}' > x'_{md.}$ if $x'_{mo.} < x'_{md.}$, and $\bar{x}' < x'_{md.}$ if $x'_{mo.} > x'_{md.}$, which are the relations given in Section IIa, for case II.

A TEST OF THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEANS OF SAMPLES FROM TWO NORMAL POPULATIONS WITHOUT ASSUMING EQUAL VARIANCES[1]

By Daisy M. Starkey

1. History of the problem. If the only available evidence about two normally distributed populations is contained in two samples, one from each, it has hitherto been the custom to the test the hypothesis that the means are equal by assuming that the quantity $\frac{\bar{x} - \bar{x}'}{\sqrt{ks^2 + k's'^2}}$ is distributed in Student's distribution, with $N + N' - 2$ degrees of freedom, where $s^2 = \frac{\Sigma(x - \bar{x})^2}{N(N - 1)}$ and $k = \frac{(N - 1)(N + N')}{N'(N + N' - 2)}$, the other notation being that used by R. A. Fisher.[2] The hypothesis underlying this test, however, is that the variances are equal. Although in many cases this may seem a reasonable assumption to adopt concurrently with that of equal means, it is undoubtedly not a necessary one, and it is, therefore, desirable that the test should be adapted to meet this difficulty.

The first advance on the problem was made by W. V. Behrens[3] who suggested that the distribution of the difference of the means could be expressed in terms of the observations in the samples from the two populations, the argument being entirely independent of the variances. R. A. Fisher[4] obtained substantially the same result, but expressed the argument in terms of fiducial probability. M. S. Bartlett[5] was of the opinion that Behrens' reasoning was incorrect, as he obtained some results which were apparently inconsistent with those tabulated in Behrens' paper, but R. A. Fisher[6] showed that Bartlett's argument was open to criticism. In the latter work, he actually obtained distributions for the case of two samples of two observations, and in the following we shall indicate some extensions of this more detailed work of Fisher, firstly, to the case

---

[1] Presented at the joint meeting of the American Mathematical Society and the Institute of Mathematical Statistics, Indianapolis, December 30, 1937.

Research done under a grant-in-aid from the Carnegie Corporation of New York City.

[2] *Statistical Methods for Research Workers*, 1925–1930.

[3] "Ein Beitrag zur Fehlerberechnung bei wenige Beobachtungen," *Landw. Jb.* 68, 807–37 (1929).

[4] "The Fiducial Argument in Statistical Inference," *Annals of Eugenics*, 6 (1935) pp. 391–8.

[5] "The Information Available in Small Samples," *Proc. Camb. Phil. Soc.* 32, pp. 560–6 (1936).

[6] "On a Point Raised by M. S. Bartlett on Fiducial Probability," *Annals of Eugenics* 7. Part IV, 370–5 (1937).

of other small samples of even numbers of observations, and, secondly, to samples of very large numbers.

2. The case of small samples. We recapitulate, briefly, the preliminary argument of R. A. Fisher[4], in which he denotes the unknown population means by μ and μ'. Since $\frac{\bar{x} - \mu}{s} = t$, where t is distributed in Student's distribution, we may write $\mu = \bar{x} - st$, and obtain the fiducial distribution of the population parameter μ in the form $G_1(\mu)\, d\mu$, where

$$G_1(\mu)\, d\mu = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \frac{d\mu}{s\left\{1 + \left(\frac{\bar{x} - \mu}{s}\right)^2 \Big/ n\right\}^{\frac{n+1}{2}}},$$

and a similar result for the fiducial distribution of μ'. The simultaneous fiducial distribution of μ and μ' is thus $G_1(\mu)\, G_2(\mu')d\mu d\mu'$ from which the fiducial distribution of $\mu - \mu'$ may be found. We may note that the characteristic function of $-(\mu - \mu') + (\bar{x} - \bar{x}')$ is $M(x)$, where

$$\begin{aligned} M(x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ix[-(\mu-\mu')+(\bar{x}-\bar{x}')]}\, G_1(\mu)G_2(\mu')\, d\mu\, d\mu' \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ix(ts-t's')}\, H_1(t)H_2(t')\, dt\, dt', \end{aligned}$$

where

$$H_1(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \frac{dt}{(1 + t^2/n)^{\frac{n+1}{2}}},$$

with a similar expression for $H_2(t)$. Thus from the fiducial point of view, the problem is essentially that of formally determining the distribution of the variate $ts - t's'$, or $at + bt'$, where $a = s$, $b = -s'$ are regarded as constants, t and t' being distributed in "Student's" distribution. The hypothesis $\mu = \mu'$ may then be examined by testing whether $\bar{x} - \bar{x}'$ is a significantly large value of this variate. We shall approach this distribution problem through the use of characteristic functions.

By definition, the characteristic function of "Student's" distribution is represented by the integral

$$(1) \qquad \frac{1}{\sqrt{\pi n}} \frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma(\frac{1}{2}n)} \int_{-\infty}^{\infty} \frac{e^{itx}}{\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}\, dt,$$

and may be evaluated by three methods which will be briefly considered.

First, by integrating the function

$$\frac{1}{\sqrt{\pi n}}\,\frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma(\frac{1}{2}n)}\,\frac{e^{it|z|}}{\left(1+\frac{z^2}{n}\right)^{\frac{n+1}{2}}}$$

around a standard semicircular contour in the upper half of the z-plane, the value of the characteristic function is at once proved to be $2\pi i$ times the sum of the residues of the integrand within the contour when the radius of the semicircle becomes infinite. Within the contour there is one pole only at $z = i\sqrt{n}$. The residue at this pole is the coefficient of $1/h$ in the expansion of

$$\frac{1}{\sqrt{\pi n}}\,\frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma(\frac{1}{2}n)}\,\frac{e^{-|t|\sqrt{n}+i|t|h}}{\left[\left(\frac{h}{\sqrt{n}}\right)\left(2i+\frac{h}{\sqrt{n}}\right)\right]^{\frac{n+1}{2}}}$$

in ascending powers of h, which may readily be evaluated when n is odd.

Second, by using the result that

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{ixt}}{1+t^2}\,dt = e^{-|x|},$$

from which we deduce that

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{ixt}}{a+t^2}\,dt = \frac{1}{\sqrt{a}}\,e^{-\sqrt{a}|x|}.$$

Differentiating this result $(n-1)$ times with respect to a, again considering odd values of n, we have that

$$(-1)^{\frac{1}{2}(n-1)}\,\frac{\left(\frac{n-1}{2}\right)!}{\pi}\int_{-\infty}^{\infty}\frac{e^{ixt}}{(a+t^2)^{\frac{n+1}{2}}}\,dt = \frac{d^{\frac{n-1}{2}}}{da^{\frac{n-1}{2}}}\left[\frac{1}{\sqrt{a}}\,e^{-\sqrt{a}|x|}\right].$$

By forming the first order differential equation in $y = \frac{1}{\sqrt{a}}\,e^{-\sqrt{a}|x|}$, and differentiating it $\frac{1}{2}(n-3)$ times using Leibnitz[5] theorem, we may obtain a linear relation between the derivatives of order $\frac{1}{2}(n-1)$ and lower; similarly, by differentiating $\frac{1}{2}(n-5)$ times, we may obtain a linear relation between the derivatives of all orders up to and including $\frac{1}{2}(n-3)$, and continuing in this way, we obtain a set of $\frac{1}{2}(n-1)$ linear equations in the $\frac{1}{2}(n-1)$ unknown derivatives. These equations may be solved for the derivative of order $\frac{1}{2}(n-1)$ by the determinant rule. The denominator determinant is independent of x, and the numerator is $e^{-|x|\sqrt{a}}$ multiplied by a polynomial of degree $\frac{1}{2}(n-1)$ in x. Using this fact, we may specify undetermined values for the coefficients in this polynomial, and obtain relations between these values for two consecutive values of n by differ-

entiating once. The recurrence relations thus obtained may be used to verify by mathematical induction the following relation, after substituting $a = \sqrt{n}$,

$$(2)\qquad \frac{1}{\sqrt{\pi n}}\frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma(\frac{1}{2}n)}\int_{-\infty}^{\infty}\frac{e^{itx}}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}\,dt$$

$$= e^{-|x|\sqrt{n}}\left[1+|x|\sqrt{n}+\frac{(|x|\sqrt{n})^2}{2!}\frac{(n-3)}{(n-2)}+\frac{(|x|\sqrt{n})^3}{3!}\frac{(n-5)}{(n-2)}\cdots\right],$$

the coefficient of $(|x|\sqrt{n})^{2k}$ being

$$\frac{1}{(2k)!}\frac{(n-4k+1)(n-4k+3)(n-4k+5)\cdots(n-2k-1)}{(n-2)(n-4)(n-6)\cdots(n-2k)},$$

and the coefficient of $(|x|\sqrt{n})^{2k+1}$ being

$$\frac{1}{(2k+1)!}\frac{(n-4k-1)(n-4k+1)\cdots(n-2k-3)}{(n-2)(n-4)\cdots(n-2k)}.$$

This is, therefore, the value of the characteristic function, and is the same in form as the result which may be obtained by the first method. There are evidently a finite number of terms, the degree of the polynomial being $\frac{1}{2}(n-1)$.

Third, the characteristic function may be shown to satisfy the second order differential equation.

$$x\frac{d^2y}{dx^2}-(n-1)\frac{dy}{dx}-nxy=0.$$

By change of variables $y = e^{-x\sqrt{n}}v$ (we assume that x is positive, as it may be replaced by its absolute value in the integral) and $u = x\sqrt{n}$, we obtain

$$u\frac{d^2v}{du^2}-\frac{dv}{du}(n-1+2u)+(n-1)v=0.$$

Using the Frobenius method of solution in series, we obtain as one solution when n is odd the expression

$$v = 1+u+\frac{u^2}{2!}\frac{(n-3)}{(n-2)}+\frac{u^3}{3!}\frac{(n-5)}{(n-2)}+\cdots$$

and the corresponding value of y has already been proved to be the value of the characteristic function. It is probable that the corresponding solution of the differential equation would also be the value of the characteristic function when n is even. In this case, however, the indicial equation has roots differing by an integer, and the solution of the differential equation is much more complicated in form. Nevertheless, it seems possible to find a series expansion for the characteristic function of "Student's" distribution in this way whatever be the value of n.

The characteristic functions of the distributions of at and bt' may now be

readily obtained by replacing x by ax and bx in the above expression. Multiplying the characteristic functions of these two independent distributions, we obtain the characteristic function of the distribution of $at + bt'$, which is of the form

$$M(x) = e^{-|x|(|a|\sqrt{n}+|b|\sqrt{n'})}\,[1 + |x|(|a|\sqrt{n} + |b|\sqrt{n'}) + \cdots],$$

the term in brackets being a polynomial of degree $\dfrac{(n + n' - 2)}{2}$. We may now use the result that the distribution is given by the integral

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\,M(x)\,dx,$$

and so obtain the distribution of $u = at + bt'$.

A distribution so obtained would involve four constants, a, b, n and n', and a derived probability table would thus be very complicated. It may, however, be simplified firstly by considering the case of equal sample numbers, and, secondly, making the transformation

$$v = \frac{(at + bt')}{|a| + |b|}, \tag{3}$$

whence the resulting distribution involves only two constants, n, and the ratio a/b. In this case the form of the characteristic function is

$$e^{-|x|\sqrt{n}}\left[1 + |x|\sqrt{n} + \frac{(|x|\sqrt{n})^2}{2!}\cdot\frac{(a^2 + b^2)\dfrac{n-3}{n-2} + 2|ab|}{(|a| + |b|)^2} + \cdots\right], \tag{4}$$

In determining the form of the distribution, we shall encounter integrals of the form

$$(n)^{\frac{1}{2}p}\int_{-\infty}^{\infty} e^{-|x|\sqrt{n}-ixv}\,|x|^p\,dx.$$

This can be reduced to

$$n^{\frac{1}{2}p}\int_0^{\infty} e^{-x\sqrt{n}-ixv}x^p\,dx + n^{\frac{1}{2}p}\int_0^{\infty} e^{-x\sqrt{n}+ixv}x^p\,dx,$$

and, integrating by parts, or using the Gamma Function integral, we obtain as the value of this integral

$$n^{\frac{1}{2}p}\,p!\left[\frac{1}{(\sqrt{n} + iv)^{p+1}} + \frac{1}{(\sqrt{n} - iv)^{p+1}}\right].$$

Writing $v = \sqrt{n}\tan\theta$, this reduces to

$$\frac{p!}{\sqrt{n}}\,2\cos(p + 1)\,\theta\cos^{p+1}\theta,$$

The distribution is thus seen to be:—

(5) $$\frac{1}{\pi}[p_0 + p_1 \cos 2\theta + p_2 \cos\theta \cos 3\theta + \cdots + p_{n-1}\cos^{n-2}\theta \cos n\theta]\, d\theta,$$

where

$$p_0 = 1, p_1 = 1, p_2 = \frac{(a^2+b^2)\dfrac{n-3}{n-2} + 2\,|ab|}{(|a|+|b|)^2}, \ldots .$$

It is obvious that the values of the coefficients p may all be expressed in terms of the ratio $\left|\dfrac{a}{b}\right|$. Denoting this ratio by r,

$$p_2 = \frac{(r^2+1)\dfrac{n-3}{n-2} + 2r}{(r+1)^2},$$

and thus we could construct a table for the probability integral involving n, r and v only.

The process of evaluating the probability integral may be simplified by considering the term already evaluated,

$$n^{\frac{1}{2}p} \int e^{-|x|\sqrt{n} - ixv} |\, x\,|^p\, dx.$$

Integrating this expression under the integral sign with respect to v, between the limits v and ∞, the contribution to the probability integral from this term is seen to be

$$\frac{n^{\frac{1}{2}p}}{i} \int_{-\infty}^{\infty} e^{-|x|\sqrt{n} - ixv} |\, x\,|^{p-1}\, dx,$$

which on the introduction of the same transformation as before, gives the value

$$-\, 2\,(p-1)!\cos^p\theta \sin p\theta.$$

Thus, from (5), the total probability that θ should lie between $\frac{\pi}{2}$ and a given value, θ, is

(6) $$\frac{1}{\pi}\left[\frac{\pi}{2} - \theta - \cos\theta\sin\theta - \frac{p_2}{2}\cos^2\theta\sin 2\theta - \cdots - \frac{p_{n-1}}{n-1}\cos^{n-1}\theta\sin(n-1)\theta\right],$$

where $-\frac{\pi}{2} \leqq \theta \leqq \frac{\pi}{2}$.

The following summarises the results for small values of n.

1. $\qquad n = 1. \qquad \tan\theta = \dfrac{at + bt'}{|a|+|b|}.$

The results reduce to those already given by Fisher. The distribution is then simply $\frac{d\theta}{\pi}$, or Student's distribution, and is independent of a and b, and the probability integral is $\frac{1}{2} - \frac{\theta}{\pi}$.

2. $n = 3$. $\tan\theta = \dfrac{at + bt'}{\sqrt{3}\,(|a| + |b|)}$.

The distribution function is

$$\frac{d\theta}{\pi}\left[1 + \cos 2\theta + \frac{2r}{(1+r)^2}\cos\theta\cos 3\theta\right],$$

and the probability integral

$$\frac{1}{\pi}\left[\frac{\pi}{2} - \theta - \cos\theta\sin\theta - \frac{r}{(1+r)^2}\cos^2\theta\sin 2\theta\right].$$

3. $n = 5$. $\tan\theta = \dfrac{at + bt'}{\sqrt{5}(|a| + |b|)}$.

The distribution function is

$$\frac{d\theta}{\pi}\left[1 + \cos 2\theta + \frac{2}{3}\frac{(r^2 + 1 + 3r)}{(1+r)^2}\cos\theta\cos 3\theta + \frac{2r}{(1+r)^2}\cos^2\theta\cos 4\theta + \frac{8r^2}{3(1+r)^2}\cos^3\theta\cos 5\theta\right],$$

and the probability integral

$$\frac{1}{\pi}\left[\frac{\pi}{2} - \theta - \cos\theta\sin\theta - \frac{1}{3}\frac{(r^2 + 1 + 3r)}{(1+r)^2}\cos^2\theta\sin 2\theta + \frac{2r}{3(1+r)^2}\cos^3\theta\sin 3\theta - \frac{2r^2}{3(1+r)^4}\cos^4\theta\sin 4\theta\right].$$

3. **Samples of large numbers.** The foregoing method is not suitable when n and n' are large. In this case we use the asymptotic expansion of "Student's" distribution which has been worked out by R. A. Fisher[7] and is of the form,

$$f(t)\,dt = \frac{1}{\sqrt{\pi n}}\frac{\Gamma[\frac{1}{2}(n+1)]}{\Gamma(\frac{1}{2}n)}\frac{dt}{\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}} \tag{7}$$

$$\sim \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}\,dt\left(1 + \frac{P_1}{n} + \frac{P_2}{n^2} + \cdots + \frac{P_k}{n^k} + \cdots\right),$$

[7] "The Expansion of 'Student's' Distribution in Powers of n^{-1}," *Metron*, Vol. 5, no. 3 (1926), pp. 22–25.

where P_k is a polynomial of degree $4k$ in t, such that

$$P_1 = \frac{t^4 - 2t^2 - 1}{4}, \qquad P_2 = \frac{3t^8 - 28t^6 + 30t^4 + 12t^2 + 3}{96}, \quad \text{etc.}$$

The development of an asymptotic expansion for the distribution of $at + bt'$ is obtained by combining the asymptotic expansions of t and t'. The theoretical justification of the process used makes use of the following lemma:—

If $R_k(t)$ is the remainder after the first $(k + 1)$ terms in the asymptotic expansion of "Student's" distribution in descending powers of n, then $\lim_{n\to\infty} n_k \int_0^\infty |R_k(t)|\, dt = 0.$

In the proof, the symbol "lim" will be used to denote the limit as n tends to infinity of the quantity in question. Let $S_k(t)$ represent the sum of the first $(k + 1)$ terms of the above expansion. It may readily be shown that if $0 < \delta < \frac{1}{4}$,

$$\lim n^k \int_{n^\delta}^\infty f(t)\, dt = 0,$$

and hence that

$$\lim n^k \int_{n^\delta}^\infty |R_k(t)|\, dt = 0.$$

Using an expansion for the logarithm of $\dfrac{1}{\sqrt{\pi n}} \dfrac{1}{\left(1 + \dfrac{t^2}{n}\right)^{\frac{n+1}{2}}}$ and the asymptotic expansion for the logarithm of the Gamma function, the following asymptotic expansion may be obtained, $\log f(t) = -\frac{1}{2}\log 2\pi - \frac{1}{2}t^2 + w$, where

$$w = \frac{1}{4n}(t^4 + 2t^2 - 1) + \frac{1}{12n^2}(-2t^6 + 3t^4) + \cdots \tag{8}$$
$$+ \frac{G_{2p+2}}{p(p+1)n^p} + \tfrac{1}{2}T_{p+1} - \frac{n+1}{2} R_p + R'_p,$$

G_{2p+2} being a polynomial of degree $2p+2$, and

$$|R_p| < \frac{t^{2p+4}}{(p+2)n^{p+2}} = \frac{\alpha t^{2p+4}}{(p+2)n^{p+2}}, \qquad \text{where } 0 < \alpha < 1,$$

$$|T_{p+1}| = \frac{t^{2p+2}}{(p+1)n^{p+1}}$$

$$|R'_p| < \frac{2^{p+2}A}{(p+1)(p+2)n^{p+1}} = \frac{\beta 2^{p+2}A}{(p+1)(p+2)n^{p+1}}, \qquad \text{where } 0 < \beta < 1,$$

A being a constant independent of n.

Thus, using Taylor's expansion, we obtain

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \left(1 + w + \frac{w^2}{2!} + \frac{w^3}{3!} + \cdots + \frac{w^k}{k!} + \frac{w^{k+1}}{(k+1)!} e^{\theta w}\right),$$

where $0 < \theta < 1$.

Evidently $R_k(t)$ is of the form

$$(9) \qquad \frac{1}{\sqrt{2\pi}} \left[e^{-\frac{1}{2}t^2} \left(\frac{q_{k+1}}{n^{k+1}} + \frac{q_{k+2}}{n^{k+2}} + \cdots + \frac{q_{k(p+2)}}{n^{k(p+2)}} + \frac{w^{k+1}}{(k+1)!} e^{\theta w} \right) \right],$$

the quantities q being polynomials in t.

Using the moments of the normal distribution, it may readily be shown that

$$\lim n^k \int_0^{n^\delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \left(\frac{q_{k+1}}{n^{k+1}} + \frac{q_{k+2}}{n^{k+2}} + \cdots + \frac{q_{k(p+2)}}{n^{k(p+2)}} \right) dt = 0.$$

In the range of integration, when n is sufficiently large, it is evident that

$$|w| < \frac{2n^{4\delta}}{n} + \frac{2n^{6\delta}}{n^2} + \cdots + \frac{2n^{(2p+2)\delta}}{n^p} + \frac{n+1}{2} |R_p| + |R'_p| + |T_{p+1}| = O(n^{4\delta-1}) \quad \text{if } 0 < \delta < \tfrac{1}{4}.$$

Thus

$$\left| \frac{w^{k+1}}{\sqrt{2\pi}(k+1)!} e^{\theta w} e^{-\frac{1}{2}t^2} \right| < \frac{K n^{\frac{1}{2}}}{n^{1-4\delta-4k\delta+k}} e^{\frac{\theta K'}{n^{1-4\delta}}}, \quad \text{where } K \text{ and } K' \text{ are constants,}$$

and hence

$$\lim n^k \int_0^{n^\delta} |R_k(t)|\, dt < \lim \frac{K}{n^{1-5\delta-4k\delta}} e^{\frac{\theta K'}{n^{1-4\delta}}} = 0, \quad \text{if } \delta < \frac{1}{5+4k}.$$

We can also deduce the following results:

1. Since the value of the integrand is unaltered if t is replaced by $-t$, we have at once

$$\lim n^k \int_{-\infty}^0 |R_k(t)|\, dt = 0.$$

2. Using both of these results it follows that

$$\lim n^k \int_{-\infty}^{\infty} |R_k(t)|\, dt = 0.$$

Hence

3. $$\lim n^k \int_{t'}^{t} |R_k(t)|\, dt = 0,$$

where t and t' have any real values, and thus it is legitimate to integrate the asymptotic expansion of $f(t)$ term by term with respect to t between any given limits.

4. If $\phi(t)$ is a function independent of n which is bounded for all values of t, the asymptotic expansion of $f(t)\phi(t)$ in terms of n may be integrated term by term with respect to t. In particular, if $\phi(t) = e^{itx}$, an asymptotic expansion for the characteristic function of "Student's" distribution may be obtained.

We may now consider the form of the distribution of $at + bt'$, and in order to simplify the argument, the following reasoning applies to the case in which the sample numbers are equal, although a similar theory may be developed for sample numbers which are of equal orders of magnitude. We may write

$$f(t) = S_k(t) + R_k(t),$$
$$f(t') = S_k(t') + R_k(t'),$$
$$u = at + bt',$$

and hence $t' = \dfrac{u - at}{b}$. The joint distribution of u and t is therefore

$$\left[S_k(t)S_k\left(\frac{u-at}{b}\right) + R_k(t)S_k\left(\frac{u-at}{b}\right) + S_k(t)R_k\left(\frac{u-at}{b}\right) + R_k(t)R_k\left(\frac{u-at}{b}\right)\right]\frac{dt\,du}{b}.$$

The distribution of u is obtained by integrating this expression with respect to t over all the possible values of t between $-\infty$ and $+\infty$. The product $S_k(t)S_k\left(\dfrac{u-at}{b}\right)$ gives the first $k+1$ terms in the asymptotic expansion which is the product of the asymptotic expansions of $f(t)$ and $f\left(\dfrac{u-at}{b}\right)$, and a remainder $\phi(t)$, where

$$\phi(t) = e^{-\frac{1}{2}t^2-\frac{1}{2}\left(\frac{u-at}{b}\right)^2}\left(\frac{v_1}{n^{k+1}} + \frac{v_2}{n^{k+2}} + \cdots + \frac{v_k}{n^{2k}}\right), \tag{10}$$

$v_1, v_2, \cdots v_k$ being polynomials in t. Let

$$R_k'(t) = \phi(t) + R_k(t)S_k\left(\frac{u-at}{b}\right) + R_k\left(\frac{u-at}{b}\right)S_k(t) + R_k(t)R_k\left(\frac{u-at}{b}\right).$$

Using the expressions for the moments of the normal distribution, it may be shown that $\int_{-\infty}^{\infty} |\phi(t)|\,dt = O\left(\dfrac{1}{n^{k+1}}\right)$. Let the upper bound of the bounded function $S_k(t)$ for all values of n and t be B. Then

$$\int_{-\infty}^{\infty}\left|S_k\left(\frac{u-at}{b}\right)R_k(t)\right|dt < B\int_{-\infty}^{\infty}|R_k(t)|\,dt$$
$$= o(n^{-k}).$$

Similarly

$$\int_{-\infty}^{\infty} \left| S_k(t) R_k\left(\frac{u - at}{b}\right) \right| dt = o(n^{-k}).$$

and

$$\int_{-\infty}^{\infty} \left| R_k(t) R_k\left(\frac{u - at}{b}\right) \right| dt = o(n^{-k}).$$

Thus

$$\lim n^k \int_{-\infty}^{\infty} | R_k'(t) | \, dt = 0$$

and hence the distribution of u may be obtained by integrating the asymptotic expansion which is the product of the asymptotic expansions of $f(t)$ and $f\left(\frac{u - at}{b}\right)$ term by term.

In practice, it is convenient to find the distribution of

$$w = \frac{at + bt'}{\sqrt{a^2 + b^2}}. \tag{11}$$

We substitute $y = \frac{bt - at'}{\sqrt{a^2 + b^2}}$ and, using the above result, it follows that the distribution of w is given by

$$\frac{dw}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} \left\{ 1 + \frac{1}{4n} \left[\frac{(aw + by)^4}{(a^2 + b^2)^2} - 2\frac{(aw + by)^2}{a^2 + b^2} + \frac{(bw - ay)^4}{(a^2 + b^2)^2} - 2\frac{(bw - ay)^2}{(a^2 + b^2)} - 2 \right] + \cdots \right\} dy$$

which is equal to

$$\frac{dw}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} \left\{ 1 + \frac{1}{4n} \left[\frac{w^4(a^4 + b^4) + 12w^2a^2b^2 + 3(a^4 + b^4)}{(a^2 + b^2)^2} - 4 - 2w^2 \right] + \cdots \right\}. \tag{12}$$

It may be noticed that this distribution may be expressed in terms of the ratio a/b only. The probability integral may readily be obtained. There is no theoretical difficulty involved in obtaining any desired number of terms of this expansion, but they rapidly become too complicated to handle with any ease. Moreover, it is difficult to find a limit of the error committed in using any given number of terms of the series for the probability integral as an approximation to the value of this integral, as the somewhat complicated method of obtaining the series masks the form of the remainder. While it is undoubtedly true that when n is very large the distribution approaches normality, and for a somewhat lower range of values of n the first two terms of

the expansion should be taken, etc., it is difficult to forecast the number of terms which should be retained for any given value of n. In fact the same difficulty seems to exist when using the original asymptotic expansion of "Student's" distribution for the calculation of probabilities. For instance, the coefficients of the powers of t which occur in the sixth term of the asymptotic expansion of the probability integral are larger than those occurring in the fifth term, and, in consequence, in spite of the greater power of n in the denominator, for certain values of n these may contribute more to the probability integral than the previous term. We are unable to say anything about the aggregate of succeeding terms in general, and, therefore, it does not seem legitimate to drop all the terms following a term which yields a contribution beyond the limit of accuracy desired. This difficulty is even more apparent in the case in which the coefficients of the various powers of t occurring in the terms beyond the first involve also the ratio a/b, and it is probable that the different values of this ratio which are possible would lead to different numbers of terms being taken for the same value of n in order to gain the same degree of precision in the probability integral.

4. **The distributions of the test quantities which correspond to (3) and (11) for equal means, when the ratio of the variances is a known quantity.** When the ratio ϕ of the variances is given, the foregoing arguments, which are independent of the parameters specifying the distribution, may no longer be applied, for this would be information not supplied by the sample. In this case, the distributions of the test quantities which have been used take forms which depend only on the ratio of the variances, and are independent of the sample estimates of the variances.

The quantity (3), used in §2, when n was a small odd number, was $\dfrac{\bar{x}-\bar{x}'}{s+s'}=v$, where $s^2=\dfrac{\Sigma(x-\bar{x})^2}{N(N-1)}$, $s'^2=\dfrac{\Sigma(x'-\bar{x}')^2}{N(N-1)}$ and $n=N-1$. On the assumption of equal population means, the distribution of this quantity takes the form

$$(13)\qquad \frac{2\Gamma(n+\frac{1}{2})\,dv}{\Gamma\left(\frac{n}{2}\right)^2\sqrt{n\pi}\,(1+\phi)^{n-1}}\int_{-\frac{1}{\sqrt{\phi}}}^{\sqrt{\phi}}\frac{(\sqrt{\phi}z+1)^{n-1}(\sqrt{\phi}-z)^{n-1}}{\left(z^2+1+\frac{v^2}{n}\right)^{n+\frac{1}{2}}}\,dz.$$

Thus in the case $n=1$, we obtain

$$\frac{dv}{\pi(1+v)^2}\left[\frac{\sqrt{\phi}}{(v^2+1+\phi)^{\frac{1}{2}}}+\frac{1}{(\phi v^2+\phi+1)^{\frac{1}{2}}}\right],$$

which is the result given by R. A. Fisher.[6] The integral may be evaluated in terms of elementary functions for small odd integral values of n.

In §3, (11), the distribution of the statistic $w=\dfrac{\bar{x}-\bar{x}'}{\sqrt{s^2+s'^2}}$ was considered

when N was large. The exact distribution may be proved to be

$$(14)\quad \phi^{\frac{n}{2}+\frac{1}{2}} \frac{\Gamma(n+\frac{1}{2})n^n(1+\phi)^n}{\sqrt{\pi}\,(w^2\phi+n(1+\phi))^{n+\frac{1}{2}}\Gamma(n)}\,dw\cdot F\left(n+\tfrac{1}{2}, \tfrac{1}{2}n, n, \frac{n(1-\phi^2)}{nw^2\phi+n(1+\phi)}\right)$$

where F is the hypergeometric function. If $\phi = 1$, we have the limiting case in which the argument of the hypergeometric function is zero, and obtain "Student's" distribution, which is to be expected in view of the evidence stated in §1, the numbers in the samples being equal.

COLUMBIA UNIVERSITY.

SOME EFFICIENT MEASURES OF RELATIVE DISPERSION[1]

By Nilan Norris

For some time it has been known that the coefficient of variation (in the sense of the ratio of the standard deviation to the arithmetic mean) is not an efficient statistic for distributions departing materially from normality.[2] At various times there have been proposed certain supplementary estimates of relative variation, such as those involving ratios between sums and differences of upper and lower quartiles, and ratios of mean deviations to medians or to arithmetic means. Some of these have appeared in certain textbooks.[3] But there appears to have been no attempt to found their use on considerations of minimum sampling variance.

The point of departure of this paper is that of using the Method of Maximum Likelihood to derive two efficient measures of relative dispersion, together with expressions for their standard errors. These optimum estimates of true or parametric variation are the ratio of the arithmetic mean to the geometric mean (the arithmetic-geometric ratio) for Pearson Type III distributions, and the ratio of the geometric mean to the harmonic mean (the geometric-harmonic ratio) for Pearson Type V distributions. The usefulness of these measures is suggested by the generalized-mean-value-function approach to the analysis of averages, especially the theorem of inequalities among averages.[4]

---

[1] Presented before a joint meeting of the American Statistical Association and the Institute of Mathematical Statistics at Chicago, Illinois on December 28, 1936.

[2] The term "efficient statistic" is used here in the sense of R. A. Fisher, that is, of a parameter-estimate which tends towards normality of distribution with the least possible standard deviation. For a discussion of the inefficiency of certain commonly used statistics as applied to distributions departing from normality, see R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions of the Royal Society of London*, Series A, Vol. 222, 1922, pp. 332–336.

[3] See, for example, William Vernon Lovitt and Henry F. Holtzclaw, *Statistics* (Prentice-Hall, Inc., New York, 1929), p. 184; Herbert Arkin and Raymond R. Colton, *Statistical Methods* (Barnes and Noble, Inc., New York, 1935), revised ed., p. 41; and Herbert Sorenson, *Statistics for Students in Psychology and Education* (McGraw-Hill Book Company, Inc., New York, 1936), pp. 153 *f.*

[4] Nilan Norris, "Inequalities among Averages," *Annals of Mathematical Statistics*, Vol. VI, No. 1, March, 1935, pp. 27–29; and "Convexity Properties of Generalized Mean Value Functions," *Annals of Mathematical Statistics*, Vol. VIII, No. 2, June, 1937, pp. 118–120. Professor John B. Canning appears to have been the first to point out the possibility of making use of certain ratio measures of relative variation. See "The Income Concept and Certain of Its Applications," *Papers and Proceedings of the Eleventh Annual Conference of the Pacific Coast Economic Association* (Edwards Brothers, Ann Arbor, 1933), p. 64.

This theorem states that if $t_1 < t_2$, then $\phi(t_1) < \phi(t_2)$, where the unit weight or simple sample type of generalized mean value function is defined as

$$(1) \qquad \phi(t) = \left(\frac{x_1^t + x_2^t + \cdots + x_n^t}{n}\right)^{\frac{1}{t}}.$$

The x_i are restricted to positive real numbers not all equal, but t may take any real value. A necessary and sufficient condition that $\phi(-\infty) = \phi(t) = \phi(\infty)$ is the excluded trivial case that $x_1 = x_2 = \cdots = x_n$. When the x_i are not all equal, the ratios between various pairs of averages as generated by $\frac{\phi(t_2)}{\phi(t_1)}$ yield ratio measures of relative dispersion, the usefulness of which depends, in part, on their efficiency as estimates of population-characterizing constants (parameters). The arithmetic-geometric ratio may be written $\frac{\phi(1)}{\phi(0)} = \frac{A}{G}$; and the geometric-harmonic ratio may be written $\frac{\phi(0)}{\phi(-1)} = \frac{G}{H}$. In certain cases it may be of convenience to reverse the order of each of the ratios. The standard errors for the two forms which each of the ratios may assume are presented below.

The demonstration that these ratio measures of relative dispersion are 100% efficient statistics for their appropriate distributions, and the derivation of useful expressions for their respective standard errors both may be accomplished by the ordinary method of differentiating the logarithm of the likelihood.

$$\text{Let digamma of } x = F_D(x) = \frac{d}{dx}\log x!,$$

$$\text{and trigamma of } x = F_T(x) = \frac{d^2}{dx^2}\log x!$$

For Pearson Type III distributions, the frequency with which the variate x falls into the range dx is given by

$$(2) \qquad df = \frac{1}{p!}\left(\frac{x}{a}\right)^p e^{-\frac{x}{a}}\frac{dx}{a}.$$

The parameter a measures the absolute dispersion of the distribution, and the parameter p determines the general shape of the frequency curve. The relative variation may be regarded as a population parameter, θ, defined as the ratio of the population arithmetic mean to the population geometric mean. Let the logarithm of the likelihood for this distribution be represented by L, we have

$$(3) \qquad L = -n\log p! - n(p+1)\log a + \Sigma\log x_i - \frac{1}{a}\Sigma x_i,$$

where the summation is taken over the n individuals of the sample. It follows that

$$(4) \qquad \frac{\partial L}{\partial a} = -\frac{n}{a}(p+1) + \frac{1}{a^2}\Sigma x_i; \quad \text{and} \quad \frac{\partial^2 L}{\partial a^2} = \frac{n}{a^2}(p+1) - \frac{2}{a^3}\Sigma x_i.$$

When L is maximized with respect to a by equating to zero the first derivative of L with respect to a, we find

$$\frac{\Sigma x_i}{n} = a(p+1). \tag{5}$$

It also follows that

$$\begin{gathered} \frac{\partial L}{\partial p} = -nF_D(p) - n\log a + \Sigma \log x_i; \\ \frac{\partial^2 L}{\partial p^2} = -nF_T(p); \quad \text{and} \quad \frac{\partial^2 L}{\partial a\,\partial p} = \frac{\partial^2 L}{\partial p\,\partial a} = -\frac{n}{a}. \end{gathered} \tag{6}$$

When L is maximized with respect to p by equating to zero the first derivative of L with respect to p, we find

$$(\Pi x_i)^{\frac{1}{n}} = ae^{F_D(p)} \tag{7}$$

The optimum estimate $\hat{p}$ of p is therefore found from (5) and (7) to be given by the equation

$$(p+1)e^{-F_D(p)} = \frac{\Sigma x_i}{n} \Big/ (\Pi x_i)^{\frac{1}{n}} = \frac{A}{G} \tag{8}$$

But $(p+1)e^{-F_D(p)}$ is the parameter θ. Hence we find the optimum estimate of θ to be $\frac{A}{G}$, which can be expressed in terms of the generalized mean value function as $\frac{\phi(1)}{\phi(0)}$. Therefore, for distributions well graduated by a Type III curve the optimum estimate of θ, the ratio of the arithmetic mean to the geometric mean, is given by $\frac{A}{G}$.

If only p is being estimated, (a given) the variance, or square of the standard deviation of p is obtained from $\frac{\partial^2 L}{\partial p^2}$, and is $V(p) = \frac{1}{nF_T(p)}$. To a first approximation, the variance of $\frac{A}{G}$, the estimate of θ, is found from the usual relation between the variance of a function and the variance of the argument, namely

$$V[f(x)] = \left[\frac{df(x)}{dx}\right]^2 V(x). \tag{9}$$

Since

$$\frac{d}{dp}(\theta) = \theta\left[\frac{1}{p+1} - F_T(p)\right], \tag{10}$$

therefore

$$V\left(\frac{A}{G}\right) = \theta^2 \frac{\left[F_T(p) - \frac{1}{p+1}\right]^2}{nF_T(p)}, \tag{11}$$

or the standard error of $\frac{A}{G}$ is the square root of the last expression, if only p is being estimated. If it is more convenient to do so, one may reverse the terms in the ratio to obtain

$$V\left(\frac{G}{A}\right) = \theta^{-2}\frac{\left[F_T(p) - \frac{1}{p+1}\right]^2}{nF_T(p)}, \tag{12}$$

and extract the square root of the last expression to obtain the standard error of $\frac{G}{A}$.

If a and p are being estimated simultaneously, there exists the matrix of negative mean values

$$\begin{Vmatrix} -E\left(\frac{\partial^2 L}{\partial a_2}\right) & -E\left(\frac{\partial^2 L}{\partial a \partial p}\right) \\ -E\left(\frac{\partial^2 L}{\partial a \partial p}\right) & -E\left(\frac{\partial^2 L}{\partial p^2}\right) \end{Vmatrix} = \begin{Vmatrix} \frac{n}{a^2}(p+1) & \frac{n}{a} \\ \frac{n}{a} & nF_T(p) \end{Vmatrix} \tag{13}$$

from which the variance of $\frac{A}{G}$ can be computed. In fact we have

$$V(\hat{p}) = \frac{p+1}{n[(p+1)F_T(p) - 1]} = \frac{1}{n\left[F_T(p) - \frac{1}{p+1}\right]}, \tag{14}$$

and consequently

$$V\left(\frac{A}{G}\right) = \theta^2\frac{\left[F_T(p) - \frac{1}{p+1}\right]}{n}. \tag{15}$$

The standard error of $\frac{A}{G}$ is equal to the square root of the expression in (15), if both a and p are being estimated. If the terms in the ratio are reversed, one obtains

$$V\left(\frac{G}{A}\right) = \theta^{-2}\frac{\left[F_T(p) - \frac{1}{p+1}\right]}{n}. \tag{16}$$

The square root of the last expression may be taken to derive the standard error of $\frac{G}{A}$. Since the digamma and the trigamma functions have been tabulated for considerable ranges,[5] these standard error formulae, and those developed below for the Type V case should be quite useful.

[5] *British Association for the Advancement of Science: Mathematical Tables* (Office of the British Association, London, 1931), Vol. I, pp. 42-51.

For Pearson Type V distributions, the frequency with which the variate x falls into the range dx is given by

$$df = \frac{1}{p!}\left(\frac{a}{x}\right)^{p+2} e^{-\frac{a}{x}}\frac{dx}{a}. \tag{17}$$

The parameter a measures the absolute dispersion of the distribution, and the parameter p determines the general shape of the frequency curve. The relative dispersion may be regarded as a population parameter, θ', defined as the ratio of the population geometric mean to the population harmonic mean. Let the logarithm of the likelihood for this distribution be represented by L. Then

$$L = -n\log p! + n(p+1)\log a - (p+2)\,\Sigma\log x_i - a\Sigma\frac{1}{x_i}, \tag{18}$$

the summation being taken over the sample of n individuals. It follows that

$$\begin{aligned}
\frac{\partial L}{\partial p} &= -nF_D(p) + n\log a - \Sigma\log x_i;\\
\frac{\partial^2 L}{\partial p^2} &= -nF_T(p);\\
\frac{\partial L}{\partial a} &= \frac{n}{a}(p+1) - \Sigma\frac{1}{x_i};\\
\frac{\partial^2 L}{\partial a^2} &= -\frac{n}{a^2}(p+1);\\
\frac{\partial^2 L}{\partial p\,\partial a} &= \frac{\partial^2 L}{\partial a\,\partial p} = \frac{n}{a}.
\end{aligned} \tag{19}$$

Let L be maximized with respect to p to derive the geometric mean, and let L be maximized with respect to a to derive $\phi(-1)$, or H, the harmonic mean. It is clear that for the Type V distribution, the relative dispersion, as we have defined it, is the population parameter $\theta' = \dfrac{e^{F_D(p)}}{p+1}$. Therefore, if $\phi(0) = G = (\Pi x_i)^{\frac{1}{n}}$, and $\phi(-1) = H = \dfrac{1}{\frac{1}{n}\Sigma\frac{1}{x_i}}$, it follows, by an argument similar to that used in the case of a Type III curve, that the geometric-harmonic ratio, $\dfrac{G}{H}$, is an optimum estimate of the parameter θ', for distributions well graduated by the Pearson Type V curve.

If only p is being estimated, the variance of $\hat{p}$ is given by $V(\hat{p}) = \dfrac{1}{nF_T(p)}$, and

$$V\left(\frac{G}{H}\right) = \theta'^2\frac{\left[F_T(p) - \dfrac{1}{p+1}\right]^2}{nF_T(p)}, \tag{20}$$

or the standard error of $\frac{G}{H}$ is the square root of the last expression, if p alone is being estimated, a being given. If the terms in the ratio are reversed,

$$V\left(\frac{H}{G}\right) = \theta'^2 \frac{\left[F_T(p) - \frac{1}{p+1}\right]^2}{nF_T(p)}, \tag{21}$$

and the square root of the last expression yields the standard error of $\frac{H}{G}$.

If a and p are being estimated simultaneously, there exists the matrix

$$\left\| \begin{matrix} -E\left(\frac{\partial^2 L}{\partial a^2}\right) & -E\left(\frac{\partial^2 L}{\partial a \partial p}\right) \\ -E\left(\frac{\partial^2 L}{\partial a \partial p}\right) & -E\left(\frac{\partial^2 L}{\partial p^2}\right) \end{matrix} \right\| = \left\| \begin{matrix} \frac{n}{a^2}(p+1) & -\frac{n}{a} \\ -\frac{n}{a} & nF_T(p) \end{matrix} \right\| \tag{22}$$

from which the variance of $\frac{G}{A}$ can be found. In fact

$$V(\hat{p}) = \frac{1}{n\left[F_T(p) - \frac{1}{p+1}\right]} \tag{23}$$

and hence

$$V(\theta') = \frac{\theta'^2}{n}\left[F_T(p) - \frac{1}{p+1}\right]. \tag{24}$$

The standard error of $\frac{G}{H}$ is then given by the square root of the expression for $V(\theta')$. If the terms in the ratio are reversed,

$$V\left(\frac{H}{G}\right) = \theta'^{-2}\left[F_T(p) - \frac{1}{p+1}\right], \tag{25}$$

the square root of which yields the standard error of $\frac{H}{G}$.

Just as the coefficient of variation is an efficient statistic only for distributions well graduated by the normal, or Pearson Type VII curve, so also the two maximum likelihood estimates of relative dispersion herein developed are efficient only when applied to their appropriate distributions. One may expect to obtain an optimum degree of efficiency only when the arithmetic-geometric ratio is used for series well specified by the Type III function, and the geometric-harmonic ratio is used for series well specified by the Type V function.

It may be recalled that Karl Pearson proposed the use of the coefficient of variation late in the nineteenth century.[5] Since that time there appears to have been some tendency to rely on it as a measure of relative variation, regard-

[5] "Regression, Heredity, and Panmixia," *Philosophical Transactions of the Royal Society of London*, Series A, Vol. 187, 1896, p. 277. For materials pertaining to the Pearson-Thorndike controversy resulting from the latter's suggestion that the ratio of the standard deviation to the square root of the arithmetic mean is often a more suitable device than is

less of whether or not it extracts from the sample a relatively large amount of the pertinent information concerning the parent population.[7] There are several cases in which the coefficient of variation is not an optimum estimate of relative dispersion. For example, in a comparison of the true or parametric variation of the weights of humans of given age levels, the arithmetic-geometric ratio is often the appropriate statistic to use, since weights tend to be distributed according to the Pearson Type III law. Frequently the distribution of weights is very well graduated by the Type V function, if the origin is fixed at 0 in advance. Although this procedure yields a special two-parameter Type V function, the principle of using the geometric-harmonic ratio as an optimum estimate of relative dispersion is still valid. Again, in a comparison of the relative variation of the personal distribution of wealth and income in certain modern countries, the arithmetic-geometric ratio will be found to have a smaller sampling variance than that of the coefficient of variation, since the personal distribution of wealth and income in these countries tends to be in accordance with the Type III law, rather than the normal law. Similarly, the distribution of the number of trials required to obtain r successes of an event having a given probability usually follows the Type III function, and requires the use of the arithmetic-geometric ratio, if the maximum amount of the relevant information is to be extracted from the sample.

It seems clear that in practice the usefulness of the arithmetic-geometric ratio and the geometric-harmonic ratio will depend on the type of the distribution with which one is dealing, and on the extent to which added efficiency is desired. In certain cases there is doubtless room for some difference of opinion as to whether or not the degree of added efficiency achieved by the use of these maximum likelihood estimates of relative dispersion will merit departing from the use of such a time-honored statistic as the coefficient of variation. If one is interested in avoiding the assumption of normality implicit in methods customarily used in the more general problem of analysis of variance, an alternative is the use of ranks.[8] Although the efficiency of these rank-correlation methods is not always 100%, their economy of effort is sometimes a great advantage.

HUNTER COLLEGE OF THE CITY OF NEW YORK.

the coefficient of variation see Edward L. Thorndike, "Empirical Studies in the Theory of Measurement," *Archives of Psychology* (The Science Press, New York, 1907), Vol. I, No. 3, April, 1907, pp. 9–13; and *An Introduction to the Theory of Mental and Social Measurements* (Teachers College, Columbia University, New York, 1913), 2d. ed. pp. 133 *f.*, or 1st. ed., 1904, pp. 102 *f.* See also Helen M. Walker, *Studies in the History of Statistical Method* (The Williams and Wilkins Company, Baltimore, 1929), p. 178.

[7] *Cf.* Walter A. Hendricks and Kate W. Robey, "The Sampling Distribution of the Coefficient of Variation," *Annals of Mathematical Statistics*, Vol. VII, No. 4, December, 1936, pp. 129–132.

[8] Harold Hotelling and Margaret Richards Pabst, "Rank Correlation and Tests of Significance Involving No Assumption of Normality," *Annals of Mathematical Statistics*, Vol. VII, No. 1, March, 1936, pp. 29–43. See also Milton Friedman, "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *Journal of the American Statistical Association*, Vol. 32, No. 200, December, 1937, pp. 675–701.

NOTES ON THE DISTRIBUTION OF THE GEOMETRIC MEAN[1]

By Burton H. Camp

There are two transformation theorems which apply particularly well to the distribution of a product and therefore to the distribution of the geometric mean of a sample. Both are implicit in the known theory of the transformation of integrals, but it is useful to state them in forms which are especially adapted to probability theory. Several examples will be considered in which distributions of the geometric mean will be derived by using these theorems.

The first theorem may be stated as

Theorem A: *Let the point set q in an N-dimensional u-space be defined so that in q a given function of the u's, $F(u_1, u_2 \cdots u_N)$ has the property that*

$$(1) \qquad \xi \leq F < \xi + d\xi.$$

Let $\bar{q}$ be the elementary volume of the point set q defined as an N-tuple integral

$$\int_q du_1 \cdots du_N$$

taken over q, having a value of order $d\xi$. Let

$$(2) \qquad u_i = \theta(t_i), \qquad i = 1, 2, \cdots N$$

be continuous and differentiable monotonic functions of the t's with unique inverses

$$(3) \qquad t_i = \theta^{-1}(u_i).$$

Let r be the point set in t-space corresponding to q in u-space under the transformation (2) with elementary volume given by the integral

$$(4) \qquad \bar{r} = \int_r dt_1 \cdots dt_N.$$

If $J(\xi)$ is defined as $\frac{dt_1}{du_1} \cdots \frac{dt_N}{du_N}$ at a point in q for which $F = \xi$, and if, for all points in q,

$$(5) \qquad \left| \frac{dt_1}{du_1} \cdots \frac{dt_N}{du_N} - J(\xi) \right| < M \cdot d\xi.$$

When M is a constant, independent of q, then the volume $\bar{r}$, is, except for terms of order $(d\xi)^2$, given by

$$(6) \qquad \bar{q}|J(\xi)|.$$

[1] Read at a joint meeting of the American Mathematical Society and the Institute of Mathematical Statistics, Indianapolis, December 30, 1937.

The proof is immediate for we have

$$\begin{aligned}\bar{r} &= \left|\int_q \frac{dt_1}{du_1}\cdots\frac{dt_N}{du_N}\,du_1\cdots du_N\right|\\ &= \left|\int_q\left[\frac{dt_1}{du_1}\cdots\frac{dt_N}{du_N}-J(\xi)\right]du_1\cdots du_N+\int_q J(\xi)\,du_1\cdots du_N\right|\\ &\leq\left|\int_q\left[\frac{dt_1}{du_1}\cdots\frac{dt_N}{du_N}-J(\xi)\right]du_1\cdots du_N\right|+\bar{q}\cdot J(\xi).\end{aligned}$$

But, by (5), the integral in the last line has a value less than $\bar{q}\,M\cdot d\xi$, and $\bar{q}$ is of order $d\xi$. Therefore $\bar{r}$ differs from $\bar{q}|J(\xi)|$ by terms of order $(d\xi)^2$.

Let us now apply this theorem to a simple case. The volume of the set q, where $\xi \leq u_1 + \cdots + u_N < \xi + d\xi$, $u_i < a$, $i = 1, \cdots, N$, can easily be shown to be

$$\bar{q} = C(Na - \xi)^{N-1}\,d\xi.$$

Let $u_i = \log t_i$. Then it follows from the theorem that

$$\bar{r} = K\,e^{\xi}(Na - \xi)^{N-1}\,d\xi,$$

$\bar{r}$ being the volume of the point set r, where

$$\xi \leq \log\,(t_1\cdots t_N) < \xi + d\xi. \tag{7}$$

By the use of (7) one can now use the geometrical method of finding the probability distribution of the geometric mean,

$$x = (t_1\cdots t_N)^{1/N}, \tag{8}$$

of samples of N from the universe $\phi(t)\,dt$, provided that $\phi\,(t_1)\cdots\phi(t_N)$ is a continuous function of ξ. Unfortunately there do not appear to be many such ϕ functions. One that is of interest is

$$\phi(t)\,dt = kt^{2s}\,dt, \qquad 0 \leq t \leq e^a.$$

Let $D(\xi)d\xi$ represent the distribution of ξ. We have

$$\begin{aligned}D(\xi)\,d\xi &= \int_r \phi(t_1)\cdots\phi(t_N)\,dt_1\cdots dt_N = \int_r k^N(t_1,\cdots t_N)^{2s}\,dt_1\cdots dt_N\\ &= \bar{r}k^N e^{2s\xi} = Ce^{\xi+2s\xi}(Na-\xi)^{N-1}\,d\xi.\end{aligned}$$

Thence we obtain as the distribution of x:

$$f(x)\,dx = C_1x^{2sN+N-1}(a - \log x)^{N-1}\,dx.$$

The form of $f(x)$ in the special case in which $s = 0$ and ϕ is a rectangle has been found by other authors,[1] and is

$$f(x)\,dx = C_1x^{N-1}(a - \log x)^{N-1}\,dx.$$

---

[1] E.g. see S. Kullback, "An application of characteristic functions to the distribution problem of statistics," *Annals of Mathematical Statistics*, vol. 5 (1934), pp. 263–270.

The second transformation theorem to be used may be stated as.

THEOREM B: *Let $\psi(u)du$ be the probability element for a given universe and let the sample $(u_1, u_2, \cdots u_N)$ be taken. Let the statistic $\xi = \gamma(u_1, u_2, \cdots u_N)$ have the distribution $F(\xi)d\xi$. If the transformation (2), satisfying the conditions imposed on it in Theorem A be applied both to the universe and to the statistic, yielding $\varphi(t)dt$ and $\xi = g(t_1, \cdots t_N)$ respectively, then the element of distribution of ξ, as obtained from ϕ, is, as before, $F(\xi)d\xi$.*

The proof is straight forward, for the distribution of ξ, as obtained from $\psi(u)du$ is given by

$$\int_q \psi(u_1) \cdots \psi(u_N)\, du_1, \cdots du_N$$

and, as obtained from $\phi(t)dt$, it is

$$\int_r \phi(t_1) \cdots \phi(t_N)\, dt_1, \cdots dt_N$$

where q is the set in u-space where $\xi \leq \gamma < \xi + d\xi$ and r is the set in t-space where $\xi \leq g < \xi + d\xi$. It is clear that these two integrals have the same value because of the relation

$$\psi(u)\, du = \psi(\theta(t))\, \frac{d\theta(t)}{dt} \cdot dt \equiv \varphi(t)\, dt$$

and the unique correspondence between the points of q and r set up by the transformation (2), with its unique inverse (3).

This theorem is particularly well adapted to the derivation of the distribution of the geometric mean because of the simple logarithmic transformation connecting the sum and the product of N numbers, and because several distributions of the sum are already known. Two of these cases will now be presented.

EXAMPLE 1. Let x be the geometric mean (8) of the sample of N from a universe with distribution law

$$\phi(t)\, dt = \frac{(\log t)^{p-1}}{t^2\Gamma(p)}\, dt \qquad (t > 1). \tag{9}$$

Then the distribution of x is

$$f(x)\, dx = \frac{N^{Np}(\log x)^{Np-1}}{x^{N+1}\Gamma(Np)}\, dx \qquad (x > 1), \tag{10}$$

and it is to be noticed that x has the same type of distribution as t.

To prove (10), first let $\xi = (u_1 + \cdots + u_N)/N$, where the u's are a sample from a Type III universe,

$$\psi(u)\, du = \frac{e^{-u}u^{p-1}}{\Gamma(p)}\, du \qquad (u > 0).$$

Irwin[3] has shown that the distribution of ξ is

$$F(\xi)\,d\xi = \frac{Ne^{-N\xi}(N\xi)^{Np-1}}{\Gamma(Np)}\,d\xi. \tag{11}$$

Making the transformation $u = \log t$, we have

$$\xi = \log(t_1 \cdots t_N)^{1/N}, \qquad \phi(t)\,dt = \frac{(\log t)^{p-1}}{t^2\Gamma(p)}\,dt, \qquad t > 1,$$

and $F(\xi)d\xi$ is unchanged. We now obtain $f(x)dx$ by substituting $\xi = \log x$ in (11).

EXAMPLE[4] 2. If x is the geometric mean (8) of a sample of N from a universe whose distribution is

$$\phi(t)\,dt = \frac{1}{tc\sqrt{2\pi}}\,e^{-\frac{1}{2c^2}\left(\log\frac{t}{G}\right)^2}\,dt, \qquad (c, t, G > 0), \tag{12}$$

the distribution of x is

$$f(x)\,dx = \frac{\sqrt{N}}{xc\sqrt{2\pi}}\,e^{-\frac{N}{2c^2}\left(\log\frac{x}{G}\right)^2}\,dx, \qquad (x > 0). \tag{13}$$

To prove this, one begins with the arithmetic mean ξ and the universe,

$$\psi(u)\,du = \frac{1}{c\sqrt{2\pi}}\,e^{-\frac{1}{2c^2}(u-\bar{u})^2}\,du. \quad \text{Here } F(\xi)\,d\xi = \frac{\sqrt{N}}{c\sqrt{2\pi}}\,e^{-\frac{N}{2c^2}(\xi-\bar{u})^2}\,d\xi.$$

Again using $u = \log t$, one obtains $\xi = \log(t_1, \cdots, t_N)^{1/N}$ and

$$\phi(t)\,dt = \frac{1}{tc\sqrt{2\pi}}\,e^{-\frac{1}{2c^2}\left(\log\frac{t}{G}\right)^2}\,dt, \text{ where } G = e^{\bar{u}} > 0,$$

and $F(\xi)\,d\xi$ is unchanged. To get (13) one substitutes $\xi = \log x$ in $F(\xi)\,d\xi$.

Again it follows that the geometric mean has the same distribution as the universe except for a change in one of the parameters (c). This frequency curve has other interesting features. It was developed by Galton and McAlister[5] by quite a different method and was called the curve of equal facility. They were seeking for a distribution $\phi(t)$ which would have the characteristic that, if t and t' were two observations differing from G by the same relative amount, $(G - t)/t = (t' - G)/G$, they would have equal probabilities. McAlister noted various properties of ϕ, including the fact that G was actually its geometric mean, and that it was not the same as the mode or the arithmetic mean. Certain properties which he did not mention are the following:

(i) If one draws a sample from a universe with the distribution ϕ in order to

---

[3] *Biometrika*, vol. 19 (1927), p. 229; see also A. Church, *Biometrika*, vol. 18 (1926), p. 386.

[4] This distribution can also be obtained by the method of A. T. Craig, *American Journal of Mathematics*, vol. 54 (1932), p. 362, but it would be difficult to evaluate his integral without the substitution which would be suggested if the distribution were known.

[5] *Proceedings of the Royal Society*, vol. 29 (1879), pp. 365, 367.

determine G, the geometric mean of the universe, the maximum likelihood solution is x, the geometric mean of the sample.

(ii) The modal point of the sampling distribution (f) approaches G as a limit as N becomes infinite.

(iii) One can devise a function s of the sample analogous to but different from Student's s, and show that x/s has a distribution independent of the parameters of G and c of the universe. To do this it is necessary first to extend the second transformation theorem so as to include cases where the number of statistics (functions of the sample) being obtained simultaneously is greater than one. This is not difficult, but since the analogous tests for significance have been developed for the normal universe it would not be particularly useful, for if the observations are distributed in accordance with $\phi(t)$ their logarithms are distributed normally, and their logarithms can equally well be used for testing significance.

(iv) If one uses the curve of equal facility instead of the normal curve as the distribution of biological lengths, then any power of such lengths, in particular the third power, which is supposed to be approximately proportional to weights, would also be distributed in the same manner, except for a change in the parameters. This is a property which the normal curve does not have. It raises the question: Can biological lengths be represented by the curve of equal facility? The remainder of this paper will be devoted to a discussion of this question and cognate matters.

The curve of equal facility may be made to approach as a limit the normal curve if the origin be moved indefinitely to the left. This is almost intuitively evident from a consideration of the hypotheses under which the two curves were derived by Galton and McAlister. It is also indicated by the behavior of the lower moments. Let ν_i refer to the ith moment of (12) relative to the origin of t, μ_i to the corresponding moment relative to the arithmetic mean. It is easy to show that

$$(14)\qquad \nu_i = G^i e^{\frac{1}{2}i^2c^2}, \quad i = 0, 1, \cdots, \qquad \nu_1 = \bar{t} = Gh, \text{ where } h = e^{\frac{1}{2}c^2},$$

$$(15)\qquad \mu_2 = G^2h^2(h^2 - 1), \qquad \mu_3 = G^3h^3(h^6 - 3h^2 + 2),$$
$$\mu_4 = G^4h^4(h^{12} - 4h^6 + 6h^2 + 3),$$

$$(16)\qquad \begin{cases} \alpha_3 = \mu_3/\mu_2^{3/2} = (h^2 + 2)\,(h^2 - 1)^{1/2}, \\ \alpha_4 = \mu_4/\mu_2^2 \;= (h^2 - 1)^4 + 6(h^2 - 1)^3 + 15(h^2 - 1)^2 + 16(h^2 - 1) + 3. \end{cases}$$

From (16) it follows that as h approaches unity α_3 and α_4 approach their normal values, 0 and 3, respectively. If at the same time μ_2 is kept constant, it follows from (15) that G^2 and therefore $\bar{t}$ become infinite. So the origin is moved an infinite distance to the left.

The question, then, whether the curve of equal facility may be used equally well with the normal curve to represent biological lengths depends on whether in practical cases the natural choice of origin, which is the position indicated by

zero length, is such as to make the two curves practically indistinguishable. This is apparently the situation in the case of human statures. For 8585 adult males born in the British Isles[6] the values of the several constants, obtained by so fitting $\phi(t)$ to the observations that the mean and standard deviations agree, are as follows: $\bar{t} = 67.46$ in., $G = 67.411$, $\sigma = 2.56$, $h = 1.00072$, observed $\alpha_3 = 0.0125$, α_3 for $\phi = 0.11$; observed $\alpha_4 = 3.149$, α_4 for $\phi = 3.02$. Thus for the curve of equal facility α_3 is further from the observed value than for the normal curve, but α_4 is nearer to its observed value. In both cases the difference is unimportant. A graph of both curves[7] would not make it clear to the eye which of the two fitted the data better.

It would be expected that the distribution of the cubes of these statures, being roughly proportional to the weights of the men, would not be normally distributed. This also can be verified easily, for the distribution of $(y = t^k)$ from $\phi(t)dt$ is $\phi(y)dy$ except that ck replaces c, and G^k replaces G. So the distribution of cubes is:

$$F(y)\, dy = \frac{1}{3cy\sqrt{2\pi}}\, e^{-\frac{1}{18c^2}\left(\log \frac{y}{G^3}\right)^2}\, dy.$$

If this curve is fitted to the cubes of the statures, $\alpha_3 = 0.23$, and $\alpha_4 = 3.21$. Both are considerably further from their normal values than before. For this case the corresponding value of h is 1.0064. It is the closeness of this quantity to unity, or in other words the smallness of the coefficient of variation, 100 $\sigma/\bar{t} = 100\,(h^2 - 1)^{1/2}$, which determines how close the curve is to the normal. For the statures $\sigma/\bar{t} = 0.0379$. For the cubes of the statures[8] $\sigma/\bar{t} = 0.269$. Its values in certain other cases[9] are: length of forearm 0.05, chest circumference 0.08, strength of grip 0.26, visual acuity 0.39. It appears to be evident, therefore, that for many types of biometric measurements, especially lengths, which we know can be represented well by the normal curve, the curve of equal facility is practically just as good. In a given case it may fit a little better or a little worse. If we wish the distribution of the arithmetic mean as obtained by sampling from such data we may find it by supposing the universe normal; if we wish the distribution of the geometric mean we may find it by supposing the universe of a curve of equal facility. This device of substituting for the normal curve another type of curve which is equally good in practical cases, in order to find the distribution of a statistic which cannot be found easily for the normal curve, may perhaps be useful also for other statistics than the geometric mean.

WESLEYAN UNIVERSITY.

---

[6] G. Udny Yule and M. G. Kendall, *An Introduction to the Theory of Statistics*, London, 1937, pp. 94, 116, 157, 163, 187.

[7] Such as on page 187, Yule and Kendall.

[8] For the weights of a similar group of men $\sigma/\bar{t} = 0.137$, and thus the two curves would be more nearly alike if fitted to weights than if fitted to the cubes of these statures.

[9] From a long list with values ranging from 0.0049 to 0.5058, compiled by Raymond Pearl, *Medical Biometry and Statistics*, Philadelphia (1930), pp. 347–9.

NOTE ON A FORMULA FOR THE MULTIPLE CORRELATION COEFFICIENT

By H. M. Bacon

There are many useful formulas available for the calculation of the multiple correlation coefficient in a k variable problem.[1] Since it frequently happens that the regression equation is the primary object of the statistical analysis, the well known formula

$$r^2_{1\cdot 23\cdots k} = \beta_{12\cdot 34\cdots k} r_{12} + \beta_{13\cdot 24\cdots k} r_{13} + \cdots + \beta_{1k\cdot 23\cdots (k-1)} r_{1k}$$

can be used to considerable advantage. While many different demonstrations of this formula are perfectly familiar, the one given in this note may prove of some interest.

First let us recapitulate briefly certain facts about the regression coefficients and the multiple correlation coefficient. Suppose we have k sets of N numbers each:

$$\begin{array}{cccc} X_{11} & X_{12} & \cdot\ \cdot & X_{1N} \\ X_{21} & X_{22} & \cdot\ \cdot & X_{2N} \\ \cdot & \cdot & \cdot\ \cdot & \cdot \\ X_{k1} & X_{k2} & \cdot\ \cdot & X_{kN}. \end{array}$$

Let $\bar{x}_j$ be the mean of the j-th set, and let $x_{ji} = X_{ji} - \bar{x}_j$. We then have k sets of N deviations from means, and we shall suppose the following k sets to be linearly independent:

$$\begin{array}{cccc} x_{11} & x_{12} & \cdot\ \cdot & x_{1N} \\ x_{21} & x_{22} & \cdot\ \cdot & x_{2N} \\ \cdot & \cdot & \cdot\ \cdot & \cdot \\ x_{k1} & x_{k2} & \cdot\ \cdot & x_{kN}. \end{array}$$

We shall consider only the regression of the "variable" x_1 upon $x_2, x_3, \cdots, x_k$. Clearly the results obtained can be made to describe the regression of any one of the variables upon the other $k - 1$ variables by rearranging the subscripts. As usual let $\lambda_2, \lambda_3, \cdots, \lambda_k$ have values which will make the sum of squares

$$F(\lambda_2, \lambda_3, \cdots, \lambda_k) = \Sigma(x_{1i} - \lambda_2 x_{2i} - \lambda_3 x_{3i} - \cdots - \lambda_k x_{ki})^2$$

a minimum. For simplicity we shall omit stating limits of summation and understand hereafter that Σ means "sum for i from $i = 1$ to $i = N$." Neces-

[1] For example, see W. J. Kirkham, "Note on the Derivation of the Multiple Correlation Coefficient", *The Annals of Mathematical Statistics*, Volume VIII (1937), pp. 68-71.

sary conditions (which are easily shown to be sufficient) are that $\lambda_2, \lambda_3, \cdots, \lambda_k$ must satisfy the equations

$$
\begin{aligned}
&\frac{\partial F}{\partial \lambda_2} = -2\Sigma(x_{1i} - \lambda_2 x_{2i} - \lambda_3 x_{3i} - \cdots - \lambda_k x_{ki})x_{2i} = 0 \\
(1)\qquad &\frac{\partial F}{\partial \lambda_3} = -2\Sigma(x_{1i} - \lambda_2 x_{2i} - \lambda_3 x_{3i} - \cdots - \lambda_k x_{ki})x_{3i} = 0 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\frac{\partial F}{\partial \lambda_k} = -2\Sigma(x_{1i} - \lambda_2 x_{2i} - \lambda_3 x_{3i} - \cdots - \lambda_k x_{ki})x_{ki} = 0.
\end{aligned}
$$

These equations are simply the "normal equations" for determining the regression coefficients. Solving them we obtain

$$
\begin{aligned}
\lambda_2 &= b_{12.34\cdots k} \\
\lambda_3 &= b_{13.24\cdots k} \\
&\cdots\cdots \\
\lambda_k &= b_{1k.23\cdots(k-1)} .
\end{aligned}
$$

The equation of regression of x_1 on $x_2, x_3, \cdots, x_k$ is therefore

$$x_1 = b_{12.34\cdots k}x_2 + b_{13.24\cdots k}x_3 + \cdots + b_{1k.23\cdots(k-1)}x_k .$$

If we let

$$u_i = b_{12.34\cdots k}x_{2i} + b_{13.24\cdots k}x_{3i} + \cdots + b_{1k.23\cdots(k-1)}x_{ki}$$

for $i = 1, 2, \cdots, N$ then $x_{1i} - u_i$ is the residual of the i-th x_1. The coefficient of multiple correlation of x_1 in terms of $x_2, x_3, \cdots, x_k$ is defined to be the simple correlation coefficient of the x's and u's:

$$r_{1.234\cdots k} = \frac{\Sigma x_{1i}u_i}{\sqrt{\Sigma x_{1i}^2 \Sigma u_i^2}}.$$

In case it is desired to express the x's in terms of their standard deviations, the following equation is used:

$$\frac{x_1}{\sigma_1} = \beta_{12.34\cdots k}\frac{x_2}{\sigma_2} + \beta_{13.24\cdots k}\frac{x_3}{\sigma_3} + \cdots + \beta_{1k.23\cdots(k-1)}\frac{x_k}{\sigma_k}$$

or

$$z_1 = \beta_{12.34\cdots k}z_2 + \beta_{13.24\cdots k}z_3 + \cdots + \beta_{1k.23\cdots(k-1)}z_k$$

where

$$
\begin{aligned}
&\beta_{12.34\cdots k} = b_{12.34\cdots k}\frac{\sigma_2}{\sigma_1} \\
(2)\qquad &\beta_{13.24\cdots k} = b_{13.24\cdots k}\frac{\sigma_3}{\sigma_1} \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\beta_{1k.23\cdots(k-1)} = b_{1k.23\cdots(k-1)}\frac{\sigma_k}{\sigma_1}
\end{aligned}
$$

and

$$z_j = \frac{x_j}{\sigma_j}.$$

Now if $\Sigma A_i B_i = 0$, *the set of numbers* $A_1, A_2, \cdots, A_N$ *is said to be orthogonal* to the set of numbers $B_1, B_2, \cdots, B_N$. Hence the conditions of equations (1) may be described by saying that the values of $\lambda_2, \lambda_3, \cdots, \lambda_k$ must be such that the set of residuals $x_{1i} - u_i$ is orthogonal to each of the $k - 1$ sets of numbers $x_{2i}, x_{3i}, \cdots, x_{ki}$. But if the set of residuals is orthogonal to each of these sets, it is orthogonal to any linear combination of them. Since the set of u's is such a linear combination, we have

$$\Sigma(x_{1i} - u_i)u_i = 0$$

and hence

$$\Sigma x_{1i} u_i = \Sigma u_i^2. \tag{3}$$

Since $u_i = b_{12.34\ldots k} x_{2i} + b_{13.24\ldots k} x_{3i} + \cdots + b_{1k.23\ldots(k-1)} x_{ki}$ it follows at once by multiplying both sides by x_{1i} and summing that

$$\Sigma x_{1i} u_i = b_{12.34\ldots k} \Sigma x_{1i} x_{2i} + b_{13.24\ldots k} \Sigma x_{1i} x_{3i} + \cdots + b_{1k.23\ldots(k-1)} \Sigma x_{1i} x_{ki}. \tag{4}$$

Writing

$$\Sigma x_{1i} x_{2i} = N\sigma_1\sigma_2 r_{12}$$

$$\Sigma x_{1i} x_{3i} = N\sigma_1\sigma_3 r_{13}$$

$$\cdots\cdots\cdots\cdots$$

$$\Sigma x_{1i} x_{ki} = N\sigma_1\sigma_k r_{1k},$$

noting the relations between the b's and the β's expressed in equations (2), and observing that we may write

$$\Sigma x_{1i} u_i = \frac{(\Sigma x_{1i} u_i)^2}{\Sigma x_{1i} u_i} = \frac{(\Sigma x_{1i} u_i)^2}{\Sigma u_i^2}$$

because of equation (3), we may therefore rewrite equation (4) as follows

$$\frac{(\Sigma x_{1i} u_i)^2}{\Sigma u_i^2} = \beta_{12.34\ldots k} \frac{\sigma_1}{\sigma_2} N\sigma_1\sigma_2 r_{12} + \beta_{13.24\ldots k} \frac{\sigma_1}{\sigma_3} N\sigma_1\sigma_3 r_{13} + \cdots$$

$$\cdots + \beta_{1k.23\ldots(k-1)} \frac{\sigma_1}{\sigma_k} N\sigma_1\sigma_k r_{1k}.$$

Now divide both sides by $\Sigma x_{1i}^2 = N\sigma_1^2$ obtaining

$$r_{1.234\ldots k}^2 = \frac{(\Sigma x_{1i} u_i)^2}{\Sigma x_{1i}^2 \Sigma u_i^2} = \beta_{12.34\ldots k} r_{12} + \beta_{13.24\ldots k} r_{13} + \cdots + \beta_{1k.23\ldots(k-1)} r_{1k}.$$

This is the formula which was to be established.

STANFORD UNIVERSITY.

THE ANNALS *of* MATHEMATICAL STATISTICS

THE OFFICIAL JOURNAL OF THE INSTITUTE
OF MATHEMATICAL STATISTICS

Contents

Vol. IX, No. 4 — December, 1938

NOTES ON HOTELLING'S GENERALIZED T

BY P. L. HSU

1. Frequency Distribution When the Hypothesis Tested is Not True

a. THE PROBLEM. Let the simultaneous elementary probability law of the $k(f+1)$ variables z_i and z'_{ir} $(i = 1, 2, \cdots, k; r = 1, 2, \cdots, f)$ be

$$(1) \quad p(z, z') = (\sqrt{2\pi})^{-k(f+1)} |C|^{\frac{1}{2}(f+1)} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} c_{ij}\{(z_i - \zeta_i)(z_j - \zeta_j) + v'_{ij}\}\right],$$

where

$$v'_{ij} = \sum_{r=1}^{f} z'_{ir} z'_{jr} \qquad (i, j = 1, 2, \cdots, k)$$

C stands for the matrix $\| c_{ij} \|$ and $|C|$, the corresponding determinant. It is required to find the elementary probability law of the statistic

$$T = |V'|^{-1} \sum_{i,j=1}^{k} V'_{ij} z_i z_j,$$

where $|V'| = |v'_{ij}|$ and V'_{ij} denotes the cofactor of the element v'_{ij} in the matrix $\| v'_{ij} \|$.

The quantity fT is a generalization of "Student's" t considered by Hotelling [1]*. It is an appropriate criterion to test the hypothesis, say H_0, that the ζ_i in the parent population as given by (1) all vanish. The distribution of T when the hypothesis H_0 is true has already been obtained by Hotelling. But our knowledge of the test is hardly complete unless we know also the distribution of T when the ζ_i do not all vanish. Indeed, only such a knowledge can enable us to control the risk of error of the second kind, i.e. of failure to detect that H_0 is untrue [3, 4].

b. THE SOLUTION. Let H be a $k \times k$ non-singular matrix such that $H'CH = I$, the unit matrix, where H' denotes the transposed matrix of H. Let the sets of variables $(z_1, z_2, \cdots, z_k)$ and $(z'_{1r}, z'_{2r}, \cdots, z'_{kr})(r = 1, 2, \cdots f)$ be subject to the same collineation by means of H, so that

$$\| z_1, z_2, \cdots, z_k \| = \| t_1, t_2, \cdots, t_k \| \cdot H'$$

$$\| z'_{1r}, z'_{2r}, \cdots, z'_{kr} \| = \| t'_{1r}, t'_{2r}, \cdots, t'_{kr} \| \cdot H' \quad (r = 1, 2, \cdots, f)$$

where the t_i and t'_{ir} are the new variables. Let further the quantities τ_i be defined by

* References are given at the end of the paper.

$$(2) \qquad \| \zeta_1, \zeta_2, \cdots, \zeta_k \| = \| \tau_1, \tau_2, \cdots, \tau_k \| \cdot H'.$$

Then, as is easy to verify, the simultaneous distribution of the new variables will be given by

$$(3) \qquad p_1(t, t') = (\sqrt{2\pi})^{-k(f+1)} \exp\left[-\frac{1}{2} \sum_{i=1}^{k} \{(t_i - \tau_i)^2 + u'_{ii}\} \right],$$

while the statistic T, as a function of the t's, retains the original form:

$$(4) \qquad T = | U' |^{-1} \sum_{i,j=1}^{k} U'_{ij} t_i t_j$$

where

$$u'_{ij} = \sum_{r=1}^{f} t'_{ir} t'_{jr} \qquad (i, j = 1, 2, \cdots, k),$$

$| U' | = | u'_{ij} |$, and U'_{ij} is the cofactor of the element u'_{ij} in the matrix $\| u'_{ij} \|$. By virtue of (2) we have the following relation between the old and new parametric constants:

$$(5) \qquad \sum_{i=1}^{k} \tau_i^2 = \sum_{i,j=1}^{k} c_{ij} \zeta_i \zeta_j.$$

Our problem is thus reduced to finding the derived distribution of T defined by (4) from the parent population given by (3).

We solve this problem by obtaining an expression for the Laplace integral $E(e^{-\beta T})$, i.e. the mathematical expectation of $e^{-\beta T}$ for real non-negative β. A few words are perhaps needed to explain the fact that the Laplace transform of an elementary probability law determines the latter uniquely except on a null set of points. If $f(x)$ is an elementary probability law which vanishes for all negative x and if

$$g(\beta) = \int_0^\infty e^{-\beta x} f(x) \cdot dx \quad \text{for} \quad \beta \geq 0,$$

then, letting c be any fixed positive constant, we have

$$g(c - \beta) = \int_0^\infty e^{\beta x} e^{-cx} f(x)\, dx$$

for all $\beta \leq c$. We get therefore

$$m_h = \int_0^\infty x^h e^{-cx} f(x)\, dx = \frac{d^h}{d\beta^h} g(c - \beta) \bigg|_{\beta=0}, \quad (h = 0, 1, 2, \cdots)$$

the definite integral being obviously finite for all $h \geq 0$. Now a sufficient condition that the set of numbers m_h determines the function $e^{-cx} f(x)$ uniquely, with the exception of a null set at most, is that the latter multiplied by $e^{k\sqrt{x}}$ be summable $(0, \infty)$ for some positive k (cf. [6], p. 320). Since this condition is trivially satisfied by the function $e^{-cx} f(x)$, this function, and therefore $f(x)$ itself,

must be uniquely determined by the m_k. In other words, $f(x)$ is uniquely determined by its Laplace transform $g(\theta)$. We now proceed to find the Laplace integral $E(e^{-\beta T})$.

Introduce the function

$$g(t, t', \theta, \alpha) = (\sqrt{2\pi})^k \, | U' |^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left\{\sum_{i,j=1}^{k} u'_{ij}\theta_i\theta_j + 2i\alpha \sum_{i=1}^{k} t_i\theta_i\right\}\right]$$

and write

$$F(t, t', \theta, \alpha) = p_1(t, t')g(t, t', \theta, \alpha),$$

where all the arguments take real values only. For any functions $\varphi(\theta)$ and $\psi(t, t')$ let us write

$$\int \varphi(\theta)\, d\theta = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(\theta)\, d\theta_1 \cdots d\theta_k$$

$$\int \psi(t, t')\, d(t, t') = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \psi(t, t')\, dt_1 \cdots dt_k\, dt'_{11} \cdots dt'_{kf}.$$

We have

$$\int d(t, t') \int | F(t, t', \theta, \alpha) |\, d\theta = \int p_1(t, t')\, d(t, t') \int g(t, t', \theta, 0)\, d\theta = 1$$

whence we know that

$$\int d(t, t') \int F\, d\theta = \int d\theta \int F\, d(t, t') \tag{6}$$

On the right-hand side of (6) we find

$$\int p_1(t, t')\, d(t, t') \int g(t, t', \theta, \alpha)\, d\theta = \int e^{-\frac{1}{2}\alpha^2 T} p_1(t, t')\, d(t, t') = E(e^{-\frac{1}{2}\alpha^2 T})$$

while for the integral on the right-hand side of (6) we have

$$\int F\, d(t, t')$$

$$= (\sqrt{2\pi})^{-k(f+2)} \exp\left(-\frac{1}{2}\sum_{i=1}^{k} \tau_i^2\right) \int \exp\left[-\frac{1}{2}\sum_{i=1}^{k} \{t_i^2 + 2(i\alpha\theta_i - \tau_i)\}\right] dt \tag{7}$$

$$\times \int | U' |^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} (\theta_i\theta_j + \delta_{ij})\, u'_{ij}\right] dt',$$

where we mean by the δ_{ij} the quantities

$$\left.\begin{aligned} \delta_{ij} &= 0 \quad \text{for } i \neq j \\ \delta_{ii} &= 1 \end{aligned}\right\} (i, j = 1, 2, \cdots, k)$$

In the equation (7) the integral with respect to the t_i is immediately written down as

$$(\sqrt{2\pi})^k \exp\left[\frac{1}{2}\sum_{i=1}^{k}(\tau_i - i\alpha\theta_i)^2\right]$$

As to the integral with respect to the t'_{ir}, we may evaluate it by the method by which Wilks [7] evaluated the moments of the generalized variance. The result is

$$2^{\frac{1}{2}k}(\sqrt{2\pi})^{kf}\,|\,\theta_i\theta_j + \delta_{ij}\,|^{-\frac{1}{2}(f+1)}\,\frac{\Gamma\frac{1}{2}(f+1)}{\Gamma(\frac{1}{2}(f+1-k))}$$

Making the substitution into (7) we get, after necessary reductions,

$$\int F\,d(t, t') = \frac{(\sqrt{\pi})^{-k}\Gamma(\frac{1}{2}(f+1))}{\Gamma(\frac{1}{2}(f+1-k))}\,|\,\theta_i\theta_j + \delta_{ij}\,|^{-\frac{1}{2}(f+1)}$$
$$\times \exp\left[-\sum_{i=1}^{k}\{\tfrac{1}{2}\alpha^2\theta_i^2 + i\alpha\tau_i\theta_i\}\right]$$

whence, noticing that $|\,\theta_i\theta_j + \delta_{ij}\,| = 1 + \sum_{i=1}^{k}\theta_i^2$

$$E(e^{-\frac{1}{2}\alpha^2 T}) = \frac{(\sqrt{\pi})^{-k}\Gamma(\frac{1}{2}(f+1))}{\Gamma(\frac{1}{2}(f+1-k))}\int\left(1+\sum_{i=1}^{k}\theta_i^2\right)^{-\frac{1}{2}(f+1)} \exp\left[-\sum_{i=1}^{k}\{\tfrac{1}{2}\alpha^2\theta_i^2 + i\alpha\tau_i\theta_i\}\right]d\theta \tag{8}$$

Equation (8) gives the Laplace transform of the elementary probability law, $p(T)$, of T. There is no essential difficulty in getting $p(T)$ by inversion directly from (8). Nevertheless, it may be of interest to get $p(T)$ indirectly by identifying the right-hand side of (8) with the Laplace transform of another elementary probability law which is otherwise *known*. For this purpose consider the simultaneous elementary probability law

$$p(x, y) = (\sqrt{2\pi})^{-(f_1+f_2)}\exp\left[-\frac{1}{2}\sum_{i=1}^{f_1}(x_i - \xi_i)^2 - \frac{1}{2}\sum_{j=1}^{f_2}y_j^2\right]$$

and let us find the derived distribution of the statistic

$$L = \sum_{i=1}^{f_1}x_i^2 \Big/ \sum_{j=1}^{f_2}y_j^2$$

As before, we introduce the function

$$g(x, y, \theta, \alpha) = (\sqrt{2\pi})^{-f_1}\left(\sum_{j=1}^{f_2}y_j^2\right)^{\frac{1}{2}f_1}\exp\left[-\frac{1}{2}\left(\sum_{j=1}^{f_2}y_j^2\sum_{i=1}^{f_1}\theta_i^2 + 2i\alpha\sum_{i=1}^{f_1}x_i\theta_i\right)\right]$$

write

$$F(x, y, \theta, \alpha) = p(x, y)g(x, y, \theta, \alpha)$$

and ascertain that

$$\int d(x, y) \int F\, d\theta = \int d\theta \int F\, d(x, y) \tag{9}$$

On the left-hand side of (9) we find

$$\int e^{-\frac{1}{2}\alpha^2 L} p(x, y)\, d(x, y) = E(e^{-\frac{1}{2}\alpha^2 L})$$

while for the integral on the right-hand side of (9), we have

$$\begin{aligned}
\int F\, d(x, y) &= (\sqrt{2\pi})^{-(2f_1+f_2)} \exp\left(-\frac{1}{2}\sum_{i=1}^{f_1} \xi_i^2\right) \\
&\quad \times \int \exp\left[-\frac{1}{2}\sum_{i=1}^{f_1} \{x_i^2 + 2(i\alpha\theta_i - \xi_i)x_i\}\right] dx \\
&\quad \times \int \left(\sum_{i=1}^{f_2} y_i^2\right)^{\frac{1}{2}f_1} \exp\left[-\frac{1}{2}\left(1 + \sum_{i=1}^{f_1} \theta_i^2\right)\sum_{i=1}^{f_2} y_i^2\right] dy \\
&= \frac{(\sqrt{\pi})^{-f_1}\Gamma(\frac{1}{2}(f_1 + f_2))}{\Gamma(\frac{1}{2}f_2)} \left(1 + \sum_{i=1}^{f_1} \theta_i^2\right)^{-\frac{1}{2}(f_1+f_2)} \\
&\quad \exp\left[-\frac{1}{2}\sum_{i=1}^{f_1} (\alpha^2\theta_i^2 + 2i\alpha\xi_i\theta_i)\right]
\end{aligned}$$

Writing

$$f_1 = k, \qquad f_2 = f + 1 - k \tag{10}$$

we get finally

$$\begin{aligned}
E(e^{-\frac{1}{2}\alpha^2 L}) &= \frac{(\sqrt{\pi})^{-k}\Gamma(\frac{1}{2}(f + 1))}{\Gamma(\frac{1}{2}(f + 1 - k))} \int \left(1 + \sum_{i=1}^{k} \theta_i^2\right)^{-\frac{1}{2}(f+1)} \\
&\quad \exp\left[-\frac{1}{2}\sum_{i=1}^{k} (\alpha^2\theta_i^2 + 2i\alpha\xi_i\theta_i)\right]
\end{aligned} \tag{11}$$

From the identity of (8) and (11) we conclude that T is distributed exactly the same as L with the appropriate "degrees of freedom" f_1 and f_2 given by (10). But the elementary probability law of L has already been derived by P. C. Tang [5]. Using his result we immediately write down the elementary probability law of T:

$$p(T) = e^{-\lambda} \sum_{h=0}^{\infty} \frac{\lambda^h}{h!} \frac{1}{B(\frac{1}{2}f_1 + h, \frac{1}{2}f_2)} T^{\frac{1}{2}f_1+h-1}(1 + T)^{-\frac{1}{2}(f_1+f_2)-h} \tag{12}$$

where f_1 and f_2 are given by (10) and

$$\lambda = \frac{1}{2}\sum_{i=1}^{k} \tau_i^2 = \frac{1}{2}\sum_{i,j=1}^{k} c_{ij}\xi_i\xi_j \tag{13}$$

in accordance with (5). The tables of probability integrals prepared by Tang can, of course, be used to suit our purpose.

2. An Optimum Property of the T-Test. To any reader familiar with the Neyman-Pearson theory of testing statistical hypotheses [3, 4], the theorem stated below may be of considerable interest.

Denote by W the $k(f + 1)$-dimensional space of the z_i and $z'_{i\nu}$ and let w be any region in W which may possibly serve as a critical region for the rejection of the hypothesis H_0. Let us speak of a critical region w as belonging to the class D if w satisfies the following condition:

$$\int_w p(z, z')d(z, z') = \epsilon + \frac{\alpha}{2} \sum_{i,j=1}^{k} c_{ij}\zeta_i\zeta_j + R \tag{14}$$

where $\epsilon < 1$ is a positive constant independent of the ζ_i, c_{ij} and the region w, α is a constant depending on w only, but not on the ζ_i or c_{ij}, and where R for any given set of values of the c_{ij} is an infinitesimal of at least the third order as all the ζ_i tend to zero.

THEOREM. *Of all the regions belonging to the class D, the particular region which gives the largest possible value to the coefficient α in the equation* (14) *is the region defined by $T \geq T_\epsilon$, where T_ϵ is a constant so determined that the probability, when all ζ_i vanish, of the observed T being not less than T_ϵ is exactly ϵ.*

The significance of the theorem is clear. Every critical region belonging to the class D serves as an unbiased exact test of the hypothesis H_0, ϵ being the preassigned chance of rejecting H_0 if it is true. Further, as is seen from (14), as the ζ_i start to depart from zero, the increased chance of rejecting H_0 due to its falsehood is approximately proportional to the quantity $\Sigma c_{ij}\zeta_i\zeta_j$. The coefficient α therefore measures the power of the critical region w to detect the falsehood of H_0, at least when the departure of the ζ_i from zero is small. Our theorem asserts that in this particular sense the T-test is the most powerful of its kind.

The method of proof is very much the same as that by which Neyman and Pearson proved some of their general theorems concerning unbiased tests. However, as the present theorem has not yet been contained in their more general results, we shall give it a full proof without referring, save in one occasion, to these authors.

PROOF. Write

$$y'_{ij} + z_i z_j = s_{ij} \qquad (i, j = 1, 2, \cdots, k)$$

$$p_0(z, z') = (\sqrt{2\pi})^{-k(f+1)} C^{\frac{1}{2}(f+1)} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} c_{ij} s_{ij}\right] \tag{15}$$

and denote by $p_0(s)$ the simultaneous elementary probability law of the variables s_{ij} derived from (15). Let W_1 be the domain of all possible positions of the point $(s_{11}, s_{12}, \cdots, s_{kk})$ in the $\frac{1}{2}k(k + 1)$-dimensional space.

We know, although we omit the proof of it, that there is no elementary probability law of the variables s_{ij} other than $p_0(s)$ which has the same moments of all orders as those derived from $p_0(z, z')$. It then follows that if $g(s)$ be any summable function of the s_{ij} and if

$$\int_{W_1} \left(\prod_{i,j=1}^{k} s_{ij}^{r_{ij}} \right) g(s) p_0(s)\, ds = 0 \tag{16}$$

for all positive integers r_{ij} or zero, then we must have $g(s) \equiv 0$ except perhaps on a null set of points.

It follows therefore that the identity

$$\int_{W_1} g(s) p_0(s)\, ds \equiv 0 \tag{17}$$

implies the identity $g(s) = 0$ provided $g(s)$ does not involve the parameters c_{ij}. For, substituting for $p_0(s)$ its expression as given by Wishart [8] we shall have

$$K \int_{W_1} g(s) p_0(s)\, ds \equiv \int_{W_1} g(s)\, |S|^{\frac{1}{2}(f-k)} \exp\left[-\frac{1}{2} \sum_{i,j=1}^{k} c_{ij} s_{ij} \right] ds \equiv 0 \tag{18}$$

where $|S| = |s_{ij}|$ and K is some constant. Differentiating (18) successively with respect to the c_{ij} and dividing the results by K, we shall regain the equations (16). Hence it follows that $g(s) \equiv 0$.

This being established, let w be any region belonging to D and rewrite the equation (14), so that

$$(\sqrt{2\pi})^{-k(f+1)} C^{\frac{1}{2}(f+1)} \int_w \exp\left[-\frac{1}{2} \sum_{i,j=1}^{k} c_{ij} \{ (z_i - \zeta_i)(z_j - \zeta_j) + v'_{ij} \} \right] d(z, z') \\ = \epsilon + \frac{\alpha}{2} \sum_{i,j=1}^{k} c_{ij} \zeta_i \zeta_j + R \tag{19}$$

Setting all the ζ_i to zero in both sides of (19), we have

$$\int_w p_0(z, z')\, d(z, z') \equiv \epsilon \tag{20}$$

identically in the c_{ij}. Differentiating (19) once with respect to ζ_i and afterwards setting all the ζ_i to zero, we easily get

$$\int_w z_i p_0(z, z')\, d(z, z') \equiv 0 \qquad (i = 1, 2, \cdots k) \tag{21}$$

for all possible values of the c_{ij}.

Finally, differentiating (19) with respect to ζ_i and then to ζ_j and putting all $\zeta_i = 0$ in the result we obtain

$$\int_w \left\{ \left(\sum_{h=1}^{k} c_{ih} z_h \right) \left(\sum_{h=1}^{k} c_{jh} z_h \right) - c_{ij} \right\} p_0(z, z')\, d(z, z') \equiv \alpha c_{ij} \qquad (i, j = 1, 2, \cdots k)$$

whence, renumbering (20)

$$\sum_{h,l=1}^{k} c_{ih} c_{jl} q_{hl} \equiv \beta c_{ij} \qquad (i, j = 1, 2, \cdots, k) \tag{22}$$

in which we denote by $\beta = \alpha + \epsilon$ and

$$q_{hl} = \int_w z_h z_l p_0(z, z')\, d(z, z') \qquad (h, l = 1, 2, \cdots, k)$$

If we denote by Q the matrix of order k formed of the elements q_{hl}, we see that (22) may be written as

$$CQC \equiv \beta C,$$

whence, since C has its inverse matrix, C^{-1},

$$Q \equiv \beta C^{-1}$$

i.e.,

(23) $$q_{ij} \equiv \beta c_{ij}^{(-1)} \qquad (i, j = 1, 2, \cdots, k)$$

where $c_{ij}^{(-1)}$ denotes the element in the matrix C^{-1} which corresponds to the element c_{ij} in the matrix C.

Conditions (20), (21) and (23) are necessary for the region w to belong to the class D. They are evidently also sufficient.

Let us evaluate the integrals in (20), (21) and the q_{ij} by first evaluating the surface integrals on any surface, say $G(s)$, on which all the s_{ij} have constant values, and then integrating the results with respect to the s_{ij} over a region, say w_1, of the s_{ij} contained in W_1. Thus we may write (20), (21) and (23) in the form

(24) $$\int_{w_1} f(s)p_0(s)\, ds \equiv \epsilon, \qquad \int_{w_1} g_i(s)p_0(s)\, ds \equiv 0, \qquad \int_{w_1} \varphi_{ij}(s)p_0(s) \equiv \beta c_{ij}^{(-1)},$$
$$(i, j = 1, 2, \cdots, k),$$

where

$$f(s) = \frac{1}{p_0(s)} \int_{G(s)} p_0(z, z')\, dG(s)$$

$$g_i(s) = \frac{1}{p_0(s)} \int_{G(s)} z_i p_0(z, z')\, dG(s)$$

$$\varphi_{ij}(s) = \frac{1}{p_0(s)} \int_{G(s)} z_i z_j p_0(z, z')\, dG(s)$$

It is readily verified that the function $p_0(z, z')/p_0(s)$ is free from the parameters c_{ij}, and consequently so are the functions $f(s)$, $g_i(s)$, $\varphi_{ij}(s)$. Besides, we can extend the definition of these functions in the whole domain W_1 by assigning them the value zero outside of the region w_1. Doing this we can now write the equations (24) as

(25) $$\int_{W_1} (f(s) - \epsilon)p_0(s)\, ds \equiv 0, \qquad \int_{W_1} g_i(s)p_0(s)\, ds \equiv 0,$$
$$\int_{W_1} [\varphi_{ij}(s) - \gamma s_{ij}]\, p_0(s)\, ds \equiv 0 \qquad (i, j = 1, 2, \cdots, k)$$

where $\gamma = \dfrac{1}{f + 1}\beta$.

Now all the equations (25) are of the form (17); consequently, according to the already established result and remembering the definitions of the functions $f(s)$, $g_i(s)$ and $\varphi_{ij}(s)$, we must have

$$\int_{G(s)} p_0(z, z')\, dG(s) = \epsilon p_0(s) \tag{26}$$

$$\int_{G(s)} z_i p_0(z, z')\, dG(s) = 0 \tag{27}$$

$$\int_{G(s)} z_i z_j p_0(z, z')\, dG(s) = \gamma s_{ij} p_0(s) \tag{28}$$

in the whole domain W_1.

Hence the most general region belonging to the class D is constructed as follows. On any surface $s_{ij} = \text{const.}$ $(i, j = 1, 2, \cdots k)$ we take an areal region such that it satisfies the equations (26)–(28); we then allow the s_{ij} to vary in the whole domain W_1. Equations (28) may now be replaced by

$$\int_{G(s)} \left(\frac{z_1^2}{s_{11}} - \frac{z_i z_j}{s_{ij}}\right) p_0(z, z') = 0, \qquad (i, j = 1, 2, \cdots, k) \tag{28'}$$

Let us call w_0 the region defined by $T' \geq T'_\epsilon$. Since w_0 belongs to the class D (cf. (12)), its cross section, say $G_0(s)$, by any surface $s_{ij} = \text{const.}$ $(i, j = 1, 2, \cdots, k)$ must satisfy the equations (26), (27) and (28′). Since $\gamma = \dfrac{1}{f+1}(\alpha + \epsilon)$, all we have to prove now is that among all the areal regions $G(s)$ satisfying the equations (26), (27) and (28′) it is the region $G_0(s)$ that gives the largest possible value to $\gamma p_0(s)$. Now

$$\gamma p_0(s) = \int_{G(s)} \frac{z_1^2}{s_{11}} p_0(z, z')\, d(z, z') \tag{29}$$

and, according to a Lemma of Neyman and Pearson, [3, p. 10] the right-hand side of (29) will attain its maximum value if $G(s)$ is defined by an inequality of the form

$$\frac{z_1^2}{s_{11}} \geq \sum_{i,j=1}^{k} a_{ij}\left(\frac{z_1^2}{s_{11}} - \frac{z_i z_j}{s_{ij}}\right) + \sum_{i=1}^{k} b_i z_i + c \tag{30}$$

where the a_{ij}, b_i and c are constants so determined as to enable the region $G(s)$ to satisfy the equations (26)–(28). We shall show presently that the region $G_0(s)$ is defined by such an inequality.

The inequality $T' \geq T'_\epsilon$ may be written as

$$\frac{|v'_{ij}|}{|v'_{ij} + z_i z_j|} \leq \frac{1}{1 + T'_\epsilon}$$

i.e.

$$\frac{|s_{ij} - z_i z_j|}{|s_{ij}|} \leq \frac{1}{1 + T'_\epsilon},$$

or

$$\sum_{i,j=1}^{k} s_{ij}^{(-1)} z_i z_j \geq \frac{T_s}{1+T_s} \tag{31}$$

where $s_{ij}^{(-1)}$ denotes the (i, j)th element in the inverse matrix of $\| s_{ij} \|$. The region $G_0(s)$ is therefore defined by the same inequality (31) in which we regard the s_{ij} as constants.

If we put

$$a_{ij} = \frac{1}{k} s_{ij} s_{ij}^{(-1)}, \qquad b_i = 0, \qquad c = \frac{1}{k}\frac{1}{1+T_s} \qquad (i, j = 1, 2, \cdots, k)$$

in (30) we can easily reduce the inequality (30) into (31).

The proof is now complete.

3. Note on Applications of T. It is already known that the T-test may be used for the following purposes (a) and (b):

(a) Given a k-variate normal surface

$$p(x) = (\sqrt{2\pi})^{-k} C^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} c_{ij}(x_i - \xi_i)(x_j - \xi_j)\right]$$

with the unknown ξ_i and c_{ij}. n observations

$$(x_{1l}, x_{2l}, \cdots, x_{kl}), \qquad (l = 1, 2, \cdots, n)$$

having been made, it is required to test the hypothesis that the ξ_i have the particular values ξ_i^0 for $i = 1, 2, \cdots, k$.

Here we use the T-test with

$$\left.\begin{aligned} z_i &= \sqrt{n}(\bar{x}_i - \xi_i^0), \qquad v'_{ij} = \sum_{l=1}^{m} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j) \\ \zeta_i &= \sqrt{n}\,(\xi_i - \xi_i^0), \qquad f = n - 1 \\ \text{where} \qquad & \\ \bar{x}_i &= \frac{1}{n}\sum_{l=1}^{n} x_{il} \end{aligned}\right\} (i, j = 1, 2, \cdots, k)$$

(b) Given two k-variate normal surfaces

$$p_1(x) = (\sqrt{2\pi})^{-k} C^{\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{k} c_{ij}(x_i - \xi_i)(x_j - \xi_j)\right)$$

$$p_2(x) = (\sqrt{2\pi})^{-k} C^{\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{k} c_{ij}(y_i - \eta_i)(y_j - \eta_j)\right)$$

where the c_{ij} are common to the two surfaces while all the ξ_i, ξ_j, c_{ij} are unknown. Samples of n_1 and n_2 having been drawn respectively from the two populations, to test the hypothesis that $\xi_i = \eta_i$ for all i.

Let the samples be

$$(x_{1l}, x_{2l}, \cdots, x_{kl}), \qquad (l = 1, 2, \cdots, n_1)$$

and

$$(y_{1h}, y_{2h}, \cdots, y_{kh}), \qquad (h = 1, 2, \cdots, n_2)$$

Let

$$\bar{x}_i = \frac{1}{n_1}\sum_{l=1}^{n_1} x_{il}, \qquad \bar{y}_i = \frac{1}{n_2}\sum_{h=1}^{n_2} y_{ih} \qquad (i = 1, 2, \cdots, k)$$

We use the T-test with

$$z_i = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}(\bar{x}_i - \bar{y}_i), \qquad v'_{ij} = \sum_{l=1}^{n_1}(x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j) + \sum_{h=1}^{n_2}(y_{ih} - \bar{y}_i)(y_{jh} - \bar{y}_j)$$

$$\zeta_i = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}(\xi_i - \eta_i), \qquad f = n_1 + n_2 - 2$$

$$(i, j = 1, 2, \cdots, k)$$

A third application of T, which appears to be novel, is the following:

(c) Given a $(k + 1)$-variate normal surface

$$p(x) = (\sqrt{2\pi})^{-(k+1)} D^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k+1} d_{ij}(x_i - \xi_i)(x_j - \xi_j)\right], \qquad D = |d_{ij}|,$$

where the ξ_i and d_{ij} are all unknown. n observations

$$(x_{1l}, x_{2l}, \cdots, x_{k+1,l}) \qquad (l = 1, 2, \cdots, n)$$

having been made, to test the hypothesis that all the ξ_i are equal.

If we put

$$y_i = x_i - x_{k+1} \qquad (i = 1, 2, \cdots, k),$$

then we have a k-variate normal surface for the variables y_i.

$$p(y) = (\sqrt{2\pi})^{-k} C^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} c_{ij}(y_i - \eta_i)\right]$$

where $\eta_i = \xi_i - \xi_{k+1}$ $(i = 1, 2, \cdots, k)$. Thus the problem is reduced to testing the hypothesis that $\eta_i = 0$ for $i = 1, 2, \cdots, k$ and therefore belongs to the type (a). Write

$$y_{il} = x_{il} - x_{k+1,l} \qquad (i = 1, 2, \cdots, k; l = 1, 2, \cdots, n)$$

and

$$\bar{y}_i = \frac{1}{2}\sum_{l=1}^{n} y_{il}, \qquad (i = 1, 2, \cdots, k).$$

We use the T-test with

$$\left.\begin{array}{ll} z_i = \sqrt{n}\,\bar{y}_i, & v'_{ij} = \sum_{l=1}^{n} (y_{il} - \bar{y})(y_{jl} - \bar{y}_j) \\ \zeta_i = \sqrt{n}\,\eta_i, & f = n + 1 \end{array}\right\} (i, j = 1, 2, \cdots, k)$$

Although there are no simple expressions for the c_{ij}, there is one for the parameter $\Sigma\, c_{ij}\eta_i\eta_j$, on which alone the distribution of T depends. We have indeed

$$\sum_{i,j=1}^{k} c_{ij}\eta_i\eta_j = \frac{1}{D} \begin{vmatrix} \sigma_{11} & \cdots & \sigma_{1,\,k+1} & \xi_1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{k+1,1} & \cdots & \sigma_{k+1,\,k+1} & \xi_{k+1} & 1 \\ \xi_1 & \cdots & \xi_{k+1} & 0 & 0 \\ 1 & \cdots & 1 & 0 & 0 \end{vmatrix}$$

where

$$D = \begin{vmatrix} \sigma_{11} & \cdots & \sigma_{1,\,k+1} & 1 \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{k+1,1} & \cdots & \sigma_{k+1,\,k+1} & 1 \\ 1 & \cdots & 1 & 0 \end{vmatrix}$$

where σ_{ij} is the covariance between x_i and x_j.

Expressing T in terms of the original variables x, we have

$$T = -\frac{1}{D'} \begin{vmatrix} s_{11}\; s_{12} & \cdots & s_{1,\,k+1} & \bar{x}_1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_{k+1,1} & \cdots & s_{k+1,\,k+1} & \bar{x}_{k+1} & 1 \\ \bar{x}_1 & \cdots & \bar{x}_{k+1} & 0 & 0 \\ 1 & \cdots & 1 & 0 & 0 \end{vmatrix}$$

where

$$D' = \begin{vmatrix} s_{11} & \cdots & s_{1,\,k+1} & 1 \\ \cdots & \cdots & \cdots & \cdots \\ s_{k+1,1} & \cdots & s_{k+1,\,k+1} & 1 \\ 1 & \cdots & 1 & 0 \end{vmatrix}$$

and where

$$\bar{x}_i = \frac{1}{n}\sum_{l=1}^{n} x_{il}, \qquad s_{ij} = \frac{1}{n}\sum_{l=1}^{n} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j), \qquad (i, j = 1, 2, \cdots, k + 1)$$

Therefore T is independent of which variable has been taken as the $(k + 1)$st.

UNIVERSITY COLLEGE, LONDON.

REFERENCES

[1] H. HOTELLING, *Ann. Math. Statist.*, Vol. 2 (1931) pp. 359–378.

[2] J. NEYMAN, *Bull. Soc. Math. France*, Vol. 63 (1935) pp. 246-266.
[3] J. NEYMAN AND E. S. PEARSON, *Statist. Res. Mem.*, Vol. 1 (1936) pp. 1-37.
[4] J. NEYMAN AND E. S. PEARSON, *Statist. Res. Mem.*, Vol. 2 (1938).
[5] P. C. TANG, *Statist. Res. Mem.*, Vol. 2 (1938).
[6] E. C. TITCHMARSH, *Introduction to the Theory of Fourier Integrals.* Oxford Univ. Press (1937).
[7] S. S. WILKS, *Biometrika*, Vol. 24 (1932) pp. 471-494.
[8] J. WISHART, *Biometrika*, Vol. 20A (1928) pp. 32-52.

GENERALIZATION OF THE INEQUALITY OF MARKOFF

By A. Wald

1. **Introduction.** Denote by X a random variable and by M_r the expected value $E|X - x_0|^r$ of $|X - x_0|^r$ for any integer r where x_0 denotes a given real value. M_r is also called the absolute moment of order r about the point x_0. For any positive number d, denote by $P(-d < X - x_0 < d)$ the probability that $|X - x_0| < d$. The inequality of Markoff can be written as follows

$$P(-d < X - x_0 < d) \geq 1 - \frac{M_r}{d^r} \tag{1}$$

The inequality (1) is also called, for $r = 2$, the inequality of Tchebyscheff. The inequality (1) can be written in the following way:

$$P(-\xi\sqrt[r]{M_r} < X - x_0 < \xi\sqrt[r]{M_r}) \geq 1 - \frac{1}{\xi^r}.$$

Substituting in the above inequality s for r and $\xi \dfrac{\sqrt[r]{M_r}}{\sqrt[s]{M_s}}$ for ξ we get

$$P(-\xi\sqrt[r]{M_r} < X - x_0 < \xi\sqrt[r]{M_r}) \geq 1 - \frac{1}{\xi^s}\left(\frac{\sqrt[s]{M_s}}{\sqrt[r]{M_r}}\right)^s, \tag{2}$$

where r and s denote any integers and ξ denotes an arbitrary positive value.[1] Substituting in (2) $2k$ for s and 2 for r, we get the inequality of K. Pearson.[2] By other substitutions we get the formulae of Lurquin, Cantelli, etc.[3]

As is well known, the inequality (1) cannot be improved[4] for $d \geq \sqrt[r]{M_r}$. That is to say, to every $\epsilon > 0$ a random variable Y can be given such that

$$E|Y - x_0|^r = E|X - x_0|^r \quad \text{and} \quad P(-d < Y - x_0 < d) < 1 - \frac{M_r}{d^r} + \epsilon.$$

If the absolute moments $M_{i_1} = E(|X - x_0|^{i_1}), \cdots, M_{i_j} = E|X - x_0|^{i_j}$ of a random variable X are given (and no further data about X are known), then we shall say that a_d is the *"sharp" lower limit* of $P(-d < X - x_0 < d)$ if the following two conditions are fulfilled:

(1) For each random variable Y, for which $E|Y - x_0|^{i_1} = E|X - x_0|^{i_1}, \cdots,$ $E|Y - x_0|^{i_j} = E|X - x_0|^{i_j}$, the inequality $P(-d < Y - x_0 < d) \geq a_d$ holds.

[1] The formula (2) has been given by A. Guldberg, *Comptes Rendus*, Paris, Vol. 175, p. 679.

[2] *Biometrika*, Vol. XII (1918-1919) pp. 284-296.

[3] E. Lurquin, *Comptes Rendus*, Paris, Vol. 175, p. 681. Also Cantelli, *Rendiconti della Reale Academia dei Lincei*, 1916.

[4] See for instance, R. v. Mises, *Wahrscheinlichkeitsrechnung*, Leipzig, Vienna, Deuticke, 1931, p. 36.

(2) To each $\epsilon > 0$, a random variable Y can be given such that $E \mid Y - x_0 \mid^{i_\nu} = E \mid X - x_0 \mid^{i_\nu}$ $(\nu = 1, \cdots, j)$ and $P(-d < Y - x_0 < d) < a_d + \epsilon$.

In other words, a_d is the *limes inferior*[5] of the probabilities $P(-d < Y - x_0 < d)$ formed for all random variables Y for which the i_ν-th absolute moment about the point x_0 is equal to the i_ν-th moment of X about the point x_0 $(\nu = 1, \cdots, j)$.

PROBLEM: *The absolute moments $M_{i_1}, M_{i_2}, \cdots, M_{i_j}$ of a random variable X are given about the point x_0, where $i_1, i_2, \cdots, i_j$ denote any integers and M_{i_ν} denotes the moment of order i_ν $(\nu = 1 \cdots k)$. It is required to calculate the "sharp" lower limit of the probability $P(-d < X - x_0 < d)$ for any positive value d.*

If only a single moment M_r is given, our problem is already solved, because the inequality (1) gives us the "sharp" lower limit for $d \geq \sqrt[r]{M_r}$ and for $d < \sqrt[r]{M_r}$ the "sharp" limit is obviously equal to zero. But the case in which even two moments M_r and M_s are given has not yet been solved. The formula (2) gives us a limit for $P(-d < X - x_0 < d)$, but this limit is not "sharp," as can easily be shown.

We shall give here some results concerning the general case, and the complete solution if only two moments M_r and M_s are given. We shall see that the "sharp" limit is considerably greater than the limit given by (2).

2. Some Propositions Concerning the General Case. We shall call a random variable X non-negative if $P(X < 0) = 0$. Since the absolute moments of the non-negative random variable $Y = \mid X - x_0 \mid$ about the origin are equal to the absolute moments of X about the point x_0 and since $P(Y < d) = P(-d < X - x_0 < d)$, the following proposition holds true:

(I) *Denote by $M_{i_1}, \cdots, M_{i_j}$ the absolute moments of order $i_1, \cdots, i_j$ of a certain random variable X about the point x_0. The limes inferior of the probabilities $P(-d < Y - x_0 < d)$ is equal to the limes inferior of the probabilities $P(Z < d)$, where $P(-d < Y - x_0 < d)$ is formed for all random variables Y for which the i_ν-th absolute moment about x_0 is equal to M_{i_ν} $(\nu = 1, \cdots, j)$, and $P(Z < d)$ is formed for all non-negative random variables Z for which the i_ν-th moment about the origin is equal to M_{i_ν} $(\nu = 1, \cdots, j)$.*

On account of the proposition (I) we can restrict ourselves to the consideration of non-negative random variables and of the moments about the origin.

A random variable X for which k different values $x_1, \cdots, x_k$ exist such that the probability $p(x_i)$ of x_i $(i = 1, \cdots, k)$ is positive and $\sum_{i=1}^{k} p(x_i) = 1$, is called an *arithmetic* random variable of degree k. A random variable X will be called t-limited, if $P(-t \leq X \leq t) = 1$ We shall prove the following proposition.

(II). *Let us denote by $M_{i_1}, M_{i_2}, \cdots, M_{i_j}$ the absolute moments of order $i_1, \cdots, i_j$ of a certain non-negative random variable X, about the origin. Denote by $\Omega(k, t)$ the set of all non-negative t-limited arithmetic random variables of*

[5] The *limes inferior* of a set N of numbers is the greatest value y for which the inequality $y \leq x$ for each element x of N holds true. This is also called greatest lower bound.

degree $\leq k$, for which the i_νth moment about the origin is equal to M_{i_ν} $(\nu = 1, \cdots, j)$. $\Omega(k, t)$ is supposed to be not empty.[6] *Denote further by $a(d, k, t)$ the limes inferior of the probabilities $P(Y < d)$ formed for all random variables Y of the set $\Omega(k, t)$. Then we can say: There exists in $\Omega(k, t)$ a random variable Z for which $P(Z < d) = a(d, k, t)$. If $0 < a(d, k, t) < 1$ and Z is a random variable in $\Omega(k, t)$ for which $P(Z < d) = a(d, k, t)$, then there exist at most $j - 1$ different positive values $x_1, \cdots, x_{j-1}$ such that $x_i \neq d$, $x_i \neq t$ and the probability $p(x_i)$ of x_i is positive $(i = 1, 2, \cdots, j-1)$.*

At first we shall prove that there exists a random variable Z in $\Omega(k, t)$ such that $P(Z < d) = a(d, k, t)$. Since $a(d, k, t)$ is the *limes inferior* of $P(Y < d)$ formed for all random variables Y in $\Omega(k, t)$, there exists in $\Omega(k, t)$ a sequence $\{Z_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ of random variables, such that $\lim_{i=\infty} P(Z_i < d) = a(d, k, t)$. Arranged in ascending order of magnitude, the values of Z_i which have a positive probability are denoted by $x_{i,1}, x_{i,2}, \cdots, x_{i,k_i}$. Since Z_i is a t-limited non-negative arithmetic random variable of degree $\leq k$, we have $k_i \leq k$ and $0 \leq x_{i,r} \leq t$ $(r = 1, \cdots, k_i)$. It follows easily from this fact that there exists a subsequence $\{Z_{i_\nu}\}$ $(\nu = 1, 2, \cdots, \text{ad inf.})$ of $\{Z_i\}$ with two properties: First, that the variables $Z_{i_1}, Z_{i_2}, \cdots$ are of the same degree (say s), that is to say $k_{i_\nu} = s$ $(\nu = 1, 2, \cdots, \text{ad inf.})$; and second, that the sequence $\{x_{i_\nu,r}\}$ $(\nu = 1, 2, \cdots, \text{ad inf.})$ converges for each integer $r \leq s$. Let us denote $\lim x_{i_\nu,r}$ by x_r $(r = 1, 2, \cdots, s)$, and the probability that Z_{i_ν} takes the value $x_{i_\nu,r}$ by $p_{i_\nu,r}$. It is obvious that there exists a subsequence $\{Z_{n_\nu}\}$ $(\nu = 1, 2, \cdots \text{ad inf.})$ of $\{Z_{i_\nu}\}$ such that the sequence $\{p_{n_\nu,r}\}$ converges with increasing ν. Let us denote $\lim_{\nu=\infty} p_{n_\nu,r}$ by p_r, $(r = 1, 2, \cdots s)$. Since $p_{i_\nu,1} + \cdots + p_{i_\nu,r} = 1$, $\sum_{r=1}^{s} p_r = 1$ must hold true. We consider the random variable Z for which the probability that $Z = x_r$ is equal to p_r $(r = 1, 2, \cdots s)$ and for which no values except $x_1, \cdots, x_r$ are possible. The random variable Z is obviously an element of $\Omega(k, t)$ and $P(Z < d) = a(d, k, t)$.

Let us consider the case in which $0 < a(d, k, t) < 1$ and denote by Z a random variable of $\Omega(k, t)$ for which $P(Z < d) = a(d, k, t)$. We shall prove that there exist at most $j - 1$ different positive values $x_1, \cdots, x_{j-1}$ such that $x_i \neq d$, $x_i \neq t$ and the probability $p(x_i)$ of x_i is positive $(i = 1, 2, \cdots, j-1)$. In order to prove this statement we shall suppose that there exist j different positive points $x_1, \cdots, x_j$ such that $x_i \neq d$, $x_i \neq t$ and $p(x_i) > 0$ $(i = 1, 2, \cdots, j)$. Then we can write

$$\sum_{\nu=1}^{j} x_\nu^{i_1} p(x_\nu) = M_{i_1} - \sum x^{i_1} p(x)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\sum_{\nu=1}^{j} x_\nu^{i_j} p(x_\nu) = M_{i_j} - \sum x^{i_j} p(x),$$

[6] This is certainly the case, if we choose k and t great enough.

where the summations on the right hand sides are to be taken over all values of x which are different from $x_1, \cdots, x_j$ and for which $p(x) > 0$.

Since $P(Z < d) = a(d, k, t)$ and $0 < a(d, k, t) < 1$ by hypothesis, there exist two non-negative values b and c, such that $b < d$, $c \geq d$, $p(b) > 0$, and $p(c) > 0$.

We define a new arithmetic random variable Z' as follows: $p'(b) = p(b) - \epsilon$, $p'(c) = p(c) + \epsilon$, and for all other values $p'(x) = p(x)$, where $p'(x)$ denotes the probability that $Z' = x$, and ϵ a positive number less than $p(b)$. Z' is obviously a non-negative arithmetic variable of the same degree as Z. The moments about the origin of the order $i_1, i_2, \cdots, i_j$ of Z' will in general not be equal to the corresponding moments of Z. However this can be obtained by a small displacement of the points $x_1, \cdots, x_j$ into a system of neighboring points $\bar{x}, \cdots, \bar{x}_j$, provided that ϵ is small enough. In order to show this, we have only to prove that the functional determinant

$$\Delta = \begin{vmatrix} i_1\bar{x}_1^{i_1-1}, & \cdots, & i_1\bar{x}_j^{i_1-1} \\ i_2\bar{x}_1^{i_2-1}, & \cdots, & i_2\bar{x}_j^{i_2-1} \\ \cdots & \cdots & \cdots \\ i_j\bar{x}_1^{i_j-1}, & \cdots, & i_j\bar{x}_j^{i_j-1} \end{vmatrix} p'(x_1) \cdots p'(x_j)$$

of the functions $f_1(\bar{x}_1, \cdots, \bar{x}_j) = \sum_{\nu=1}^{j} \bar{x}_\nu^{i_1} p'(x_\nu), \cdots, f_j(\bar{x}_1, \cdots \bar{x}_j) = \sum_{\nu=1}^{j} \bar{x}_\nu^{i_j} p'(x_\nu)$ does not vanish at the point $\bar{x}_1 = x_1, \cdots, \bar{x}_j = x_j$. Since $p'(x_1), p'(x_2), \cdots p'(x_j)$ are not equal to zero, we have only to show that

$$\Delta^* = \begin{vmatrix} x_1^{i_1-1}, & \cdots, & x_j^{i_1-1} \\ \cdots & \cdots & \cdots \\ x_1^{i_j-1}, & \cdots, & x_j^{i_j-1} \end{vmatrix} = \begin{vmatrix} 1, & \cdots, & 1 \\ x_1^{i_2-i_1}, & \cdots, & x_j^{i_2-i_1}, \\ \cdots & \cdots & \cdots \\ x_1^{i_j-i_1}, & \cdots & x_j^{i_j-i_1} \end{vmatrix} x_1^{i_1-1} \cdots x_j^{i_1-1} \neq 0$$

where $i_2 - i_1, \cdots, i_j - i_1$ can be assumed positive by denoting by i_1 the smallest of the integers $i_1, i_2, \cdots, i_j$.

Let us consider the polynomial in x given by

$$R(x) = \begin{vmatrix} 1, & \cdots, & 1, & 1 \\ x_1^{i_2-i_1}, & \cdots & x_{j-1}^{i_2-i_1}, & x^{i_2-i_1} \\ \cdots & \cdots & \cdots & \cdots \\ x_1^{i_j-i_1}, & \cdots, & x_{j-1}^{i_j-i_1}, & x^{i_j-i_1} \end{vmatrix}$$

According to a well-known algebraic proposition, the number of positive roots of $R(x)$ is less than or equal to the number of changes of sign in the sequence of coefficients of $R(x)$. Since the number of changes of sign in $R(x)$ is obviously less than or equal to $j - 1$, the number of positive roots of $R(x)$ is also less than or equal to $j - 1$. On the other hand $x = x_1, \cdots, x = x_{j-1}$ are $j - 1$ positive roots of $R(x)$. Hence for any positive value $x \neq x_1, \neq x_2, \cdots,$

$\neq x_{j-1}$ the polynomial $R(x)$ does not vanish. Thus $R(x_j)$ and therefore also Δ^* and Δ are not equal to zero.

Let us denote by Z^* the random variable which we get from Z' by a small displacement of the points $x_1, \cdots, x_j$ into a system of neighboring points $\bar{x}_1, \cdots, \bar{x}_j$, such that the moment of order i_ν of Z^* about the origin becomes equal to M_{i_ν} $(\nu = 1, 2, \cdots, j)$. By choosing ϵ small enough we can obtain the values $\bar{x}_1, \cdots, \bar{x}_j$ as near to $x_1, \cdots, x_j$ as we like. In particular, ϵ can be chosen so small that $\bar{x}_1, \cdots, \bar{x}_j$ are positive numbers less than t, and $\bar{x}_i > d$ or $< d$ accordingly as $x_i >$ or $< d$. Then Z^* is obviously an element of $\Omega(k, t)$. But for Z^*

$$P(Z^* < d) = P(Z' < d) = P(Z < d) - \epsilon = a(d, k, t) - \epsilon$$

holds true, which is a contradiction because $a(d, k, t)$ is the *limes inferior* of $P(Y < d)$ formed for all random variables Y contained in $\Omega(k, t)$. Hence our assumption that there exist j different positive numbers $x_1, \cdots, x_j$, for which $x_i \neq d$, $x_i \neq t$ and $p(x_i) > 0$ $(i = 1, 2, \cdots, j)$, cannot be true, and the proposition II is proved in all its parts.

It follows from the proposition II that $a(d, k, t)$ is independent of k. On account of this fact and of the fact that any random variable X can be arbitrarily well approximated by arithmetic random variables, we get the proposition:

III. *Let us denote by $M_{i_1}, \cdots, M_{i_j}$ the moments about the origin of order $i_1, \cdots, i_j$ of a certain non-negative random variable. Denote by $\Omega(t)$ the set of all non-negative t-limited random variables, for which the i_ν-th moment about the origin is equal to M_{i_ν} $(\nu = 1, \cdots, j)$. Denote further by $a(d, t)$ the limes inferior of the probabilities $P(Y < d)$ formed for all random variables Y contained in $\Omega(t)$. Then we can say: There exists in $\Omega(t)$ a random variable Z for which $P(Z < d) = a(d, t)$. If $0 < a(d, t) < 1$ and Z is a random variable for which $P(Z < d) = a(d, t)$, then there exist at most $j - 1$ different positive numbers $x_1, \cdots, x_{j-1}$, such that $x_i \neq d$, $x_i \neq t$, and the probability that $Z = x_i$, is positive $(i = 1, 2, \cdots, j - 1)$*

It is obvious that $a(d, t)$ decreases monotonically with increasing t. Hence $\lim_{t=\infty} a(d, t)$ exists and it can be easily shown that:

$a(d, t)$ converges towards a_d if $t \to \infty$.

3. Solution of the Problem if Only Two Moments are Given. Let us denote by M_r and M_s the absolute moments respectively of order r and s about the point x_0 of a certain random variable X, where r and s $(r < s)$ denote any integers.

Let us first consider the case

$$\text{(a)} \qquad \frac{M_r}{d^r} \leq \frac{M_s}{d^s}.$$

It follows from (1) that

$$a_d \geq 1 - \frac{M_r}{d^r}$$

We shall show that $a_d = 1 - \frac{M_r}{d^r}$ if $\frac{M_r}{d^r} \leq 1$. For this purpose let us consider the arithmetic random variable Y_b of degree 3 defined as follows:

$$p(x_0 + d) = \frac{M_r}{d^r} - \frac{\epsilon}{2}, \qquad p(x_0 + d + b) = \frac{\epsilon}{2}\left(\frac{d}{d+b}\right)^r$$

$$p(x_0) = 1 - p(x_0 + d) - p(x_0 + d + b)$$

where ϵ is a positive number and $p(u)$ denotes the probability for $Y_b = u$. The r-th moment about x_0 of Y_b is obviously equal to M_r. On account of (α) the s-th moment of Y_b about x_0 is less than or equal to M_s for $b = 0$. On the other hand the s-th moment of Y_b about x_0 will be greater than M_s if b is sufficiently large. Hence there exists a non-negative value b_0 such that the s-th moment of Y_{b_0} is equal to M_s.

Since $P(-d < Y_{b_0} - x_0 < d) = 1 - \frac{M_r}{d^r} + \frac{\epsilon}{2} - \frac{\epsilon}{2}\left(\frac{d}{d+b_0}\right)^r < 1 - \frac{M_r}{d^r} + \frac{\epsilon}{2}$ and since ϵ can be chosen arbitrarily small, we have

$$a_d = 1 - \frac{M_r}{d^r}.$$

If $\frac{M_r}{d^r} \geq 1$, then a_d is equal to zero, because a_d decreases monotonically with decreasing d and $a_d = 0$ for $d = \sqrt[r]{M_r}$.

We have now to consider the case

$$(\beta) \qquad \frac{M_r}{d^r} > \frac{M_s}{d^s}.$$

First we shall show that

$$(3) \qquad \frac{M_r}{d^r} < 1.$$

In fact, if $\frac{M_r}{d^r}$ were ≥ 1, then making use of (β) we have $\left(\frac{M_r}{d^r}\right)^{\frac{s}{r}} \geq \frac{M_r}{d^r} > \frac{M_s}{d^s}$, and hence $(M_r)^{\frac{s}{r}} > M_s$. But this is not possible, because according to the well-known inequalities between moments, $(M_r)^{\frac{s}{r}}$ is less than or equal to M_s. It follows from (3) and (β) that

$$(4) \qquad \frac{M_s}{d^s} < 1.$$

In order to calculate a_d, we shall apply the propositions found in section 2. On account of the proposition I, a_d is equal to the *limes inferior* of $P(Y < d)$

where $P(Y < d)$ is formed for all non-negative random variables Y for which the r-th moment about the origin is equal to M_r and the s-th moment about the origin is equal to M_s. Hence we can restrict ourselves to the consideration of non-negative random variables and of the moments about the origin.

We shall show that $0 < a(d, t)$ holds for any positive value t. In order to prove this, it is sufficient to show that $a_d > 0$ since $a(d, t) \geq a_d$. It follows from the inequality (1) that $a_d \geq 1 - \frac{M_r}{d^r}$. Since, according to (3), $\frac{M_r}{t^r} < 1$, we have $a_d > 0$, and therefore also

(5)
$$a(d, t) > 0$$

Let us see whether $a(d, t) < 1$. If $M_s = (M_r)^{\frac{s}{r}}$, then, as is well-known, only a single non-negative random variable X exists for which the r-th moment about the origin is equal to M_r and the s-th moment is equal to $(M_r)^{\frac{s}{r}}$, namely the arithmetic random variable X of degree 1 for which the probability that $X = \sqrt[r]{M_r}$ is equal to 1. Since $\sqrt[r]{M_r} < d$, as can be seen from (3), we have $P(X < d) = 1$, and therefore $a_d = 1$. Hence in this case our problem is already solved and we have to consider only the alternative:

(6)
$$M_s = M_r^{\frac{s}{r}} + \sigma^2 \; (\sigma^2 > 0)$$

We shall show that $a(d, t) < 1$ for $t > \sqrt[r]{M_r} + d_r$. For this purpose let us consider the non-negative arithmetic random variable Y_ϵ of the degree 3 defined as follows:

$$p(\sqrt[r]{M_r}) = 1 - \epsilon, \; p(t) = \epsilon \frac{M_r}{t^r} < \epsilon \frac{M_r}{t^r} < \epsilon$$

$$p(0) = 1 - p(\sqrt[r]{M_r}) - p(t) = \epsilon - \epsilon \frac{M_r}{t^r},$$

where $p(u)$ denotes the probability for $Y_\epsilon = u$, and ϵ is a positive number < 1.

The r-th moment of Y_ϵ is equal to

$$M_r p(\sqrt[r]{M_r}) + t^r p(t) = M_r.$$

The s-th moment of Y_ϵ is given by the expression

$$A = M_r^{\frac{s}{r}} p(\sqrt[r]{M_r}) + t^s p(t) = (1 - \epsilon) M_r^{\frac{s}{r}} + \epsilon t^s \frac{M_r}{t^r}.$$

On account of (6), A is less than M_s for $\epsilon = 0$. For $\epsilon = 1$ we have

$$A = t^{s-r} M_r > d^{s-r} M_r.$$

Since from (β) $d^{s-r} M_r > M_s$, we have $A > M_s$ for $\epsilon = 1$. Hence there exists a positive value $\epsilon_0 < 1$ for which $A = M_s$. Thus the r-th moment of Y_{ϵ_0} is equal to M_r and the s-th moment of Y_{ϵ_0} is equal to M_s. We have

$$P(Y_{t_0} < d) = p(0) + p(\sqrt[r]{M_r}) = \epsilon - \epsilon\frac{M_r}{t^r} + 1 - \epsilon = 1 - \epsilon\frac{M_r}{t^r} < 1.$$

Hence

(7) $$\alpha(d, t) < 1.$$

On account of (5) and (7) it follows from proposition III, that there exists a non-negative arithmetic random variable X belonging to the set $\Omega(t)$ such that $P(X < d) = \alpha(d, t)$ and there exists at most one positive value $\delta(\neq d, \neq t)$ with positive probability. Hence $\alpha(d, t)$ is equal to the *limes inferior* of the probabilities $P(Y < d)$ formed for all non-negative arithmetic random variables Y which have the following two properties:

(1) The r-th moment about the origin is equal to M_r and the s-th moment about the origin is equal to M_s

(2) There exists at most a single positive value $\delta(\neq d, \neq t)$ with positive probability.

Denote by Z a non-negative t-limited random variable with the properties (1), (2), and for which $P(Z < d) = \alpha(d, t)$. The following equations hold

(8)
$$\begin{aligned} p(0) + p(\delta) + p(d) + p(t) &= 1 \\ p(\delta)\delta^r + p(d)d^r + p(t)t^r &= M_r \\ p(\delta)\delta^s + p(d)d^s + p(t)t^s &= M_s \end{aligned}$$

where $p(u)$ denotes the probability that $Z = u$.

From the last two equations of (8), we get

(9) $$p(\delta) = \frac{M_r d^{s-r} - M_s + p(t)\,[t^s - t^r d^{s-r}]}{\delta^r(d^{s-r} - \delta^{s-r})}$$

(10) $$p(d) = \frac{M_s - \delta^{s-r} M_r + p(t)\,[t^r\delta^{s-r} - t^s]}{d^r(d^{s-r} - \delta^{s-r})}.$$

Since $\dfrac{M_r}{d^r} > \dfrac{M_s}{d^s}$ and $t > d$, the numerator in (9) is positive. Since $0 \leq p(\delta) \leq 1$, the inequality

(11) $$0 < \delta < d$$

must hold. Hence

(12) $$p(\delta) > 0.$$

We shall show that $p(t) = 0$ if t is sufficiently large. For this purpose let us make the assumption $p(t) > 0$. We define a new random variable Z' as follows:

$$p'(t) = p(t) - \epsilon \text{ where } 0 < \epsilon < p(t)$$

$$p'(d) = p(d) - \epsilon \frac{t^r\delta^{s-r} - t^s}{d^r(d^{s-r} - \delta^{s-r})}$$

$$p'(\delta) = p(\delta) - \frac{\epsilon(t^s - t^r d^{s-r})}{\delta^r(d^{s-r} - \delta^{s-r})}$$

$$p'(0) = 1 - p'(\delta) - p'(d) - p'(t)$$

and

$$p'(z) = 0 \text{ for all values } z \neq 0, \neq \delta, \neq d, \neq t.$$

$p'(u)$ denotes the probability that $Z' = u$.

The equations (8) remain satisfied if we substitute $p'(0)$, $p'(\delta)$, $p'(d)$, and $p'(t)$ for $p(0)$, $p(\delta)$, $p(d)$, and $p(t)$ respectively. Hence the r-th moment of Z' is equal to M_r and the s-th moment is equal to M_s. We have to show that Z' is in fact a random variable, that is to say, that the defined probabilities are ≥ 0 and ≤ 1. It is sufficient to show that the defined probabilities are non-negative, because the sum of them is equal to 1 and therefore they must be ≤ 1.

Obviously $p'(t)$ is >0. Since $t > d$ and according to (11) $d > \delta$, we have $p'(d) > p(d) > 0$. According to (12), $p(\delta)$ is positive. Hence for ϵ sufficiently small $p'(\delta)$ is also positive. We have to show that also $p'(0) \geq 0$. $p'(0)$ is given by

$$\begin{aligned} p'(0) &= 1 - p'(\delta) - p'(d) - p'(t) \\ &= 1 - p(\delta) - p(d) - p(t) + \epsilon\left[1 + \frac{t^r\delta^{s-r} - t^s}{d^r(d^{s-r} - \delta^{s-r})} + \frac{t^s - t^r d^{s-r}}{\delta^r(d^{s-r} - \delta^{s-r})}\right] \\ &= p(0) + \epsilon\frac{d^r\delta^r(d^{s-r} - \delta^{s-r}) + t^s(d^r - \delta^r) - t^r(d^s - \delta^s)}{d^r\delta^r(d^{s-r} - \delta^{s-r})}, \end{aligned}$$

Since $p(0) \geq 0$, $\epsilon > 0$, $d > \delta$ and $s > r$, this last expression is positive if t is sufficiently large. We may assume t so great that $p'(0) \geq 0$, because we want to calculate only

$$a_d = \lim_{t=\infty} a(d, t).$$

Now we shall show that

$$p'(d) + p'(t) > p(d) + p(t).$$

In fact

$$\begin{aligned} p'(d) + p'(t) - p(d) - p(t) &= \epsilon\left[\frac{t^s - t^r\delta^{s-r}}{d^r(d^{s-r} - \delta^{s-r})} - 1\right] \\ &= \epsilon\left[\frac{t^r}{d^r}\frac{t^{s-r} - \delta^{s-r}}{d^{s-r} - \delta^{s-r}} - 1\right] > 0. \end{aligned}$$

Then

$$p'(0) + p'(\delta) < p(0) + p(\delta) = \alpha(d, t)$$

must follow. Since $p'(0) + p'(\delta) = P(Z' < d)$, we have a contradiction and therefore the assumption $p(t) > 0$ is reduced to an absurdity. Hence $p(t)$ must be equal to zero and $\alpha(d, t) = \alpha_d$. If we substitute zero for $p(t)$ in (8), (9), and (10) we obtain:

$$\text{(13)} \qquad \begin{cases} p(0) + p(\delta) + p(d) = 1 \\ p(\delta)\delta^r + p(d)d^r \qquad = M_r \\ p(\delta)\delta^s + p(d)d^s \qquad = M_s \end{cases}$$

$$\text{(14)} \qquad p(\delta) = \frac{M_r d^{s-r} - M_s}{\delta^r(d^{s-r} - \delta^{s-r})}$$

$$\text{(15)} \qquad p(d) = \frac{M_s - M_r \delta^{s-r}}{d^r(d^{s-r} - \delta^{s-r})}.$$

We shall prove that $p(0) = 0$. For this purpose let us make the assumption $p(0) > 0$. Denote by δ_1 a positive number $<\delta$ and let us consider the arithmetic random variable Z' of degree 3 defined as follows:

$$p'(\delta_1) = \frac{M_r d^{s-r} - M_s}{\delta_1^r(d^{s-r} - \delta_1^{s-r})}$$

$$p'(d) = \frac{M_s - M_r\delta_1^{s-r}}{d^r(d^{s-r} - \delta_1^{s-r})}$$

$$p'(0) = 1 - p'(\delta_1) - p'(d).$$

The r-th moment of Z' is evidently equal to M_r and the s-th moment to M_s. Since $p(\delta) > 0$ according to (12), and $p(0) > 0$ by hypothesis, $p'(0)$ and $p'(\delta_1)$ will be greater than zero if δ_1 is sufficiently near to δ. The derivative of $p'(d)$ with respect to δ_1 is given by

$$\frac{1}{d^r}\,\frac{-M_r(s-r)\delta_1^{s-r-1}(d^{s-r} - \delta_1^{s-r}) + (s-r)\delta_1^{s-r-1}(M_s - M_r\delta_1^{s-r})}{(d^{s-r} - \delta_1^{s-r})^2}$$

$$= \frac{(s-r)\delta_1^{s-r-1}}{d^r(d^{s-r} - \delta_1^{s-r})^2}\,(M_s - M_r d^{s-r}).$$

Since $\dfrac{M_r}{d^r} > \dfrac{M_s}{d^s}$, the above expression is negative. Hence $p'(d)$ decreases with increasing δ_1. Since $\delta_1 < \delta$, we have

$$p'(d) > p(d) \geq 0$$

and therefore

$$1 - p'(d) < 1 - p(d) = a_d .$$

Since $1 - p'(d) = P(Z' < d)$, we have a contradiction and the assumption $p(0) > 0$ is proved an absurdity. Hence $p(0) = 0$, and $p(\delta) + p(d) = 1$. From (13), (14) and (15) we have

$$q(\delta) + p(d) = \frac{M_r d^s - M_s d^r + M_s \delta^r - M_r \delta^s}{\delta^r d^r (d^{s-r} - \delta^{s-r})} = 1.$$

Hence

$$M_r d^s - M_s d^r + \delta^r (M_s - d^s) + \delta^s (d^r - M_r) = 0. \tag{16}$$

The equation (16) in δ has at most two positive roots, because the derivative of the left hand side of (16)

$$r\delta^{r-1}(M_s - d^s) + s\delta^{s-1}(d^r - M_r)$$

has exactly one positive root in δ. Since $\delta = d$ is a root of (16), the value of δ which we are seeking must be the second positive root of (16), which we shall denote by δ_0.

It can be easily shown that $\delta_0 \leq \sqrt[r]{M_r} < d$. In fact, for $\delta = 0$ the left hand side of (16) is positive on account of the assumption (β) and for $\delta = \sqrt[r]{M_r}$ it becomes equal to

$$M_s(M_r - d^r) - M_r^{\frac{s}{r}}(M_r - d^r) = \left(M_s - M_r^{\frac{s}{r}}\right)(M_r - d^r)$$

Since $M_s \geq M_r^{\frac{s}{r}}$ and recalling from (8) that $M_r < d^r$, the above expression is less than or equal to 0. Hence δ_0 lies between 0 and $\sqrt[r]{M_r} < d$.

Hence a_d is given by the expression

$$a_d = 1 - p(d) = 1 - \frac{M_s - M_r \delta_0^{s-r}}{d^r(d^{s-r} - \delta_0^{s-r})}. \tag{17}$$

For $s = 2r$ the root δ_0 can be easily calculated. We get

$$\delta_0 = \sqrt[r]{\frac{M_{2r} - d^r M_r}{M_r - d^r}} \tag{18}$$

If we substitute in (17) $2r$ for s and the right hand side of (18) for δ_0, then we get

$$\begin{aligned}
a_d &= 1 - \frac{M_{2r} - M_r\left(\dfrac{M_{2r} - d^r M_r}{M_r - d^r}\right)}{d^r\left(d^r - \dfrac{M_{2r} - d^r M_r}{M_r - d^r}\right)} \\
&= 1 - \frac{(M_r - d^r)M_{2r} - M_r(M_{2r} - d^r M_r)}{d^r[d^r(M_r - d^r) - M_{2r} + M_r d^r]} \\
&= 1 - \frac{d^r(M_r^2 - M_{2r})}{d^r[2M_r d^r - d^{2r} - M_{2r}]} \\
&= 1 - \frac{M_r^2 - M_{2r}}{2M_r d^r - d^{2r} - M_{2r}}.
\end{aligned}$$

Let us denote the non-negative number $M_{2r} - M_r^2$ by σ^2, then we obtain[7]

$$(19) \qquad a_d = 1 - \frac{\sigma^2}{(d^r - M_r)^2 + \sigma^2}. \qquad (\sigma^2 = M_{2r} - M_r^2).$$

Let us compare the "sharp" limit given by (19) with the limit given by (2). If we substitute, in (2), $2r$ for s and d for $\xi\sqrt[r]{M_r}$ we have

$$b_d = 1 - \frac{M_{2r}}{d^{2r}} = 1 - \left(\frac{M_r}{d^r}\right)^2 - \frac{\sigma^2}{d^{2r}}$$

as a lower limit of the probability $P(-d < X < x_0 < d)$. We see that for small values of σ^2, b_d is considerably smaller than a_d.

Our results may be summarized in the following

THEOREM: *Denote by M_r the r-th and by M_s the s-th absolute moment of a random variable X about the point x_0, where $r < s$. For any positive value d denote by $P(-d < X < x_0 < d)$ the probability that $|X - x_0| < d$. The "sharp" lower limit a_d of $P(-d < X - x_0 < d)$ is defined as the limes inferior of the probabilities $P(-d < Y - x_0 < d)$ formed for all random variables Y for which the r-th moment about x_0 is equal to M_r and the s-th moment about x_0 is equal to M_s. We have to distinguish two cases.*

I. $\frac{M_r}{d^r} \le \frac{M_s}{d^s}$. *In this case* $a_d = 1 - \frac{M_r}{d^r}$ *if* $\frac{M_r}{d^r} \le 1$, *and* $a_d = 0$ *if* $\frac{M_r}{d^r} > 1$.

II. $\frac{M_r}{d^r} > \frac{M_s}{d^s}$. In this case a_d is given by

$$(17) \qquad a_d = 1 - \frac{M_s - M_r\delta_0^{s-r}}{d^r(d^{s-r} - \delta_0^{s-r})},$$

where δ_0 is the positive root $\neq d$ of the equation[8] in δ

$$M_r d^s - M_s d^r + \delta^r(M_s - d^s) + \delta^s(d^r - M_r) = 0.$$

For $s = 2r$ we have

$$\delta_0 = \sqrt[r]{\frac{M_{2r} - d^r M_r}{M_r - d^r}}.$$

If we substitute in (17) $2r$ for s and the above expression for δ_0, we obtain

$$a_d = 1 - \frac{\sigma^2}{(d^r - M_r)^2 + \sigma^2},$$

where $\sigma^2 = M_{2r} - M_r^2$.

COWLES COMMISSION, COLORADO SPRINGS.

[7] The case $s = 2r$ has been treated also by Cantelli. He demonstrated the formula (19) in quite another way, which cannot be generalized for the case $s \neq 2r$. Cantelli's formula and its demonstration are given in the book of M. Frechet, *Generalities sur Probabilities. Variables Aleatoires*, Paris, 1937, pp. 123–126.

[8] As has been shown, there exists exactly one positive root $\neq d$ of the equation considered.

A MODIFICATION OF BAYES' PROBLEM

By R. v. Mises

The classical Bayes problem can be stated as follows. We consider an urn which contains white and black balls (or balls designated by 0 and 1). The probability p for drawing a black ball is unknown. But there is given a probability function $F(x)$ representing the *a priori* probability for the inequality $p \leq x$. We draw n times from the urn (returning each time the extracted ball) and get a black ball m times and a white one $n - m$ times. Now, after this experiment, we ask for the *a posteriori* probability $P_n(x)$ for the relation $p \leq x$.

The solution proposed by Bayes can be written in a slightly generalized form:

$$P_n(x) = K \int_0^x p^m(1 - p)^{n-m} dF(p) \tag{1}$$

where K is a constant to be found by means of the condition

$$P_n(1) = 1. \tag{1'}$$

We are interested in the behaviour of $P_n(x)$ if n tends to ∞ under the condition

$$\lim_{n\to\infty} \frac{m}{n} = \alpha. \tag{2}$$

Laplace found in the case of a priori equipartition $F(x) = x$, and I proved in 1919[1] for any derivable $F(x)$, that $P_n(x)$ tends to a normal distribution:

$$\lim_{n\to\infty} \left[P_n(x) - \frac{1}{\sqrt{\pi}} \int_{-\infty}^{u} e^{-u^2} du \right] = 0 \tag{3}$$

with

$$u = H_n(x - A_n)$$

$$A_n = \alpha, \qquad \frac{1}{2H_n^2} = \frac{\alpha(1 - \alpha)}{n}. \tag{4}$$

It is easily seen from (3) and (4) that

$$\lim_{n\to\infty} P_n(x) = \begin{cases} 0 & \text{if } x < \alpha \\ 1 & \text{" } x > \alpha. \end{cases} \tag{5}$$

Let us now consider a slightly modified form of the problem.[2] Instead of one

---

[1] *Mathematische Zeitschrift*, vol. 4 (1919) p. 92. Cf. my textbook *Wahrscheinlichkeitsrechnung und ihre Anwendungen*, Wien-Leipzig 1931, p. 158. Later I proved the Laplace-Bayes theorem for a more general class of $F(x)$: *Monatshefte für Mathematik und Physik*, vol. 43 (1936) pp. 105–128.

[2] This modified problem has been treated by S. Bochner, *Annals of Math.*, Vol. 37, 1936, p. 816.

urn we suppose there are given n urns each containing white and black balls. The probability p_ν for drawing a black ball from the ν^{th} urn is unknown, but is subject to an a priori probability function $F(x)$ which furnishes the a priori probability for the relation $p_\nu \leq x$, independently of ν. We assume that on drawing one ball from every urn a black ball appears m times and a white ball $n - m$ times. Putting

$$\frac{p_1 + p_2 + \cdots + p_n}{n} = p, \tag{6}$$

we ask for the a posteriori probability $P_n(x)$ for the relation $p \leq x$.

The Bayes formula (1) must now be replaced by

$$\begin{aligned} P_n(x) = K' \int\int \cdots \int_{p_1+p_2+\cdots p_n \leq nx} p_1 p_2 \cdots p_m (1 - p_{m+1})(1 - p_{m+2}) \\ \cdots (1 - p_n)\, dF(p_1) \cdots dF(p_n) \end{aligned} \tag{7}$$

where K' is a constant determined by (1'). It is very easy to examine the asymptotic character of (7). We shall prove the following

THEOREM: *If the first three moments of the a priori distribution $F(x)$*

$$b_\nu = \int_0^1 x^\nu \, dF(x), \qquad \nu = 1, 2, 3 \tag{8}$$

exist and if the dispersion $b_2 - b_1^2$ is different from 0, the a posteriori probability $P_n(x)$ tends for $n \to \infty$ under the condition (2) to the normal distribution (3) with

$$\begin{aligned} A_n &= \alpha \frac{b_2}{b_1} + (1 - \alpha) \frac{b_1 - b_2}{1 - b_1} \\ \frac{1}{2H_n^2} &= \frac{1}{n}\left[\alpha \frac{b_1 b_3 - b_2^2}{b_1^2} + (1 - \alpha) \frac{(b_1 - b_2)(1 - b_1) - (b_1 - b_2)^2}{(1 - b_1)^2} \right]. \end{aligned} \tag{9}$$

In order to prove the theorem we write

$$\begin{aligned} V_\nu(p_\nu) &= \frac{1}{b_1} \int_0^{p_\nu} x\, dF(x), \quad \text{if } \nu = 1, 2, \cdots m \\ &= \frac{1}{1 - b_1} \int_0^{p_\nu} (1 - x)\, dF(x), \quad \text{if } \nu = m + 1, m + 2, \cdots n. \end{aligned} \tag{10}$$

Then formula (7) becomes

$$P_n(x) = C \int\int \cdots \int_{p_1+p_2+\cdots p_n \leq nx} dV_1(p_1)\, dV_2(p_2) \cdots dV_n(p_n). \tag{11}$$

Each $V_\nu(p_\nu)$ is a distribution function, i.e. a non-decreasing function with $V_\nu(-\infty) = 0$, $V_\nu(\infty) = 1$. Therefore the constant C in (11) is equal to 1 and

the integral represents the distribution function for the arithmetical mean $(p_1 + p_2 + \cdots p_n)/n$. According to the *Central Limit Theorem* of the theory of probability $P_n(x)$ will converge towards a normal distribution when certain conditions are satisfied. In every case, if a_ν, s_ν^2 denote the mean value and the dispersion associated with $V_\nu(x)$, then the mean value A_n and the dispersion S_n^2 associated with $P_n(x)$ will be defined by

$$A_n = \frac{1}{n}\sum_{\nu=1}^{n} a_\nu, \qquad S_n^2 = \frac{1}{n^2}\sum_{\nu=1}^{n} s_\nu^2. \tag{12}$$

We find from (10)

$$\begin{aligned} a_\nu &= \int_0^1 x\,dV_\nu(x) = \frac{1}{b_1}\int_0^1 x^2\,dF(x) = \frac{b_2}{b_1}, \quad \text{if } \nu = 1, 2, \cdots m \\ &= \frac{1}{1-b_1}\int_0^1 x(1-x)\,dF = \frac{b_1 - b_2}{1-b_1}, \quad \text{if } \nu = m+1, \cdots n \end{aligned} \tag{13}$$

$$\begin{aligned} s_\nu^2 &= \int_0^1 x^2\,dV_\nu(x) - a_\nu^2 = \frac{b_3}{b_1} - \frac{b_2^2}{b_1^2}, \quad \text{if } \nu = 1, 2, \cdots m \\ &= \frac{b_2 - b_3}{1-b_1} - \frac{(b_1-b_2)^2}{(1-b_1)^2}, \quad \text{if } \nu = m+1, \cdots n. \end{aligned} \tag{14}$$

We supposed the dispersion of $F(x)$ to be different from zero. It follows that

$$b_1 \neq 0,\ 1 - b_1 \neq 0,\ b_3 b_1 - b_2^2 \neq 0,\ (b_2 - b_3)(1 - b_1) - (b_1 - b_2)^2 \neq 0. \tag{15}$$

For $b_1 = 0$ would imply that $dF(x) = 0$ for all $x > 0$ and $1 - b_1 = 0$ that $dF(x) = 0$ for all $x < 1$; in both cases the dispersion would be zero. On the other hand, it is easily seen that the relation $b_3 b_1 - b_2^2 = 0$ is not compatible with the condition of a non-vanishing a priori dispersion and that the same is true for the relation $(b_2 - b_3)(1 - b_1) - (b_1 - b_2)^2 = 0$.

The total dispersion Σs_ν^2 is equal to the sum of m times the value $(b_3 b_1 - b_2^2)/b_1^2$ and $n - m$ times the value $[(b_2 - b_3)(1 - b_1) - (b_1 - b_2)^2]/(1 - b_1)^2$.

Thus we see that under the condition (2) the sum Σs_ν^2 tends to ∞, while the ratio $s_\nu^2/\Sigma s_\nu^2$ tends to zero, if n increases infinitely. These are sufficient conditions for the validity of the Central Limit Theorem.[3] The values given for A_n and H_n^2 in (9) follow from (12), (13), (14) and the well known relation $2H_n^2 S_n^2 = 1$.

S. Bochner in his previously quoted paper found, in a more complicated manner, the value of A_n and only showed that $P_n(x)$ tends to zero if $x < A_n$ and to 1 if $x > A_n$.

Examples. If we assume the a priori probability to be uniform, i.e. $F(x) = x$, we have

$$b_1 = \tfrac{1}{2}, \qquad b_2 = \tfrac{1}{3}, \qquad b_3 = \tfrac{1}{4}$$

and therefore from (9)

[3] Cf. H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tract in Mathematics and Mathematical Physics, No. 36, 1937, p. 56.

$$A_n = \tfrac{1}{3}(\alpha + 1), \qquad \frac{1}{2H_n^2} = \frac{1}{18n}.$$

A more general case is that of a more concentrated a priori probability function

$$F'(x) = Cx^k(1-x)^l, \qquad C = \frac{(k+l+1)!}{k!\,l!}.$$

Here we find

$$b_1 = \frac{k+1}{k+l+2}, \qquad b_2 = \frac{(k+1)(k+2)}{(k+l+2)(k+l+3)},$$

$$b_3 = \frac{(k+1)(k+2)(k+3)}{(k+l+2)(k+l+3)(k+l+4)}$$

and the values of A_n and H_n^2 are

$$A_n = \frac{\alpha+k+1}{k+l+3}, \qquad \frac{1}{2H_n^2} = \frac{\alpha(l-k)+(k+1)(l+2)}{n(k+l+3)^2(k+l+4)}.$$

By introducing the moments of $F(x)$ relative to the mean value, i.e.

$$(16) \qquad \begin{aligned} B_2 &= \int_0^1 (x-b_1)^2\,dF = b_2 - b_1^2, \\ B_3 &= \int_0^1 (x-b_1)^3\,dF = b_3 - 3b_1b_2 + 2b_1^3 \end{aligned}$$

we can transform the general formulas (9) into

$$(17) \qquad \begin{aligned} A_n &= b_1 + \frac{B_2}{b_1(1-b_1)}(\alpha - b_1) \\ \frac{1}{2H_n^2} &= \frac{1}{n}\left[B_2 + B_3\frac{\alpha-b_1}{b_1(1-b_1)} - B_2^2\,\frac{b_1^2+\alpha(1-2b_1)}{b_1^2(1-b_1)^2}\right]. \end{aligned}$$

The first of these equations shows that the a posteriori mean value A_n (for all n) is equal to the a priori mean value b_1, if the experimental mean m/n or α coincides with the latter. On the other hand, in the case of a symmetric a priori distribution ($b_1 = \frac{1}{2}$, $B_3 = 0$) the second equation is reduced to

$$\frac{1}{2H_n^2} = \frac{1}{n}(B_2 - 4B_2^2).$$

On the whole it is remarkable that the influence of the a priori probability does not vanish for $n \to \infty$, in the case of our modified Bayes problem.[4] The explanation of this fact is to be found in a more generalized theory of the inverse problems in probability.

UNIVERSITY OF ISTANBUL, TURKEY.

[4] Cf. my papers quoted in footnote 1.

ON THE PROBABILITY THEORY OF ARBITRARILY LINKED EVENTS

BY HILDA GEIRINGER

1. Introduction. The classical Poisson problem can be stated as follows: Let $p_1, p_2, \cdots p_n$ be the probabilities of n *independent* events $E_1, E_2, \cdots E_n$ respectively; i.e. the probability of the simultaneous occurrence of E_i and E_j is equal to $p_i p_j$, that of E_i, E_j, E_k is equal to $p_i p_j p_k$ and so on. We seek the probability $P_n(x)$ that x of the events shall occur. If, $p_1 = p_2 = \cdots = p_n$ the problem is known as the Bernoulli problem.

More generally the n events may be regarded as *dependent*. Let p_{ij} be the probability of the simultaneous occurrence of E_i and E_j; p_{ijk} that of E_i, E_j, E_k and finally $p_{12\cdots n}$ that of $E_1, E_2, \cdots E_n$. There shall arise again the problem of determining the probability $P_n(x)$ that x of the n events will take place.[1] Furthermore the asymptotic behaviour of $P_n(x)$ for large n can be studied; and we shall especially be interested in the problem of the convergence of $P_n(x)$ towards a normal distribution or a Poisson distribution.

Even in the general case which we just explained, the sums

$$S_1 = \sum_{i=1}^{n} p_i, \qquad S_2 = \sum_{i,j=1}^{n} p_{ij}, \cdots S_n = p_{12\cdots n}$$

of our probabilities differ only by constant factors from the *factorial moments* $M_n^{(1)}, M_n^{(2)}, \cdots M_n^{(n)}$ of $P_n(x)$. For we have

$$S_\nu = \frac{1}{\nu!} M_n^{(\nu)} = \frac{1}{\nu!} \sum_{x=\nu}^{n} x(x-1)\cdots(x-\nu+1)P_n(x).$$

Starting from this remark the author has, in earlier papers, [8, 9, 10] established a theory of the asymptotic behaviour of $P_n(x)$, making use of the theory of moments. The criterion for the convergence of $P_n(x)$ towards the normal—or the Poisson—distribution consists of certain conditions[2] which the S_ν must satisfy.

In the following section a concise statement of the whole problem will be given, independently of the author's earlier publications. For the convergence towards the normal distribution we shall be able to establish a theorem under wider conditions in a manner which seems to be simpler. Finally, some applications of the theory will be considered.

---

[1] See, for instance, references [1]–[7] at end of paper.

[2] Using the "theorem of the continuity of moments," Professor v. Mises [11] established sufficient conditions for the convergence of $P_n(x)$ towards a Poisson distribution in the case of the problem of "iterations." However, his reasoning can be applied to the general case without much difficulty.

2. Formulation of the problem. Let us consider the n-dimensional *collective* (Kollectiv) consisting of a sequence of any n trials. In the simplest case these trials will be *alternatives*, i.e. for every trial there will exist only two results, which we may denote by "occurrence," "non-occurrence" or by "1," "0." The single trial may eventually be composed in various manners. For instance we may draw $m > n$ times from an urn, which contains counters, bearing in arbitrary proportions numbers from 0 to 9. The first "event" E_1 may consist of the fact that the first three extracted counters bear even numbers; the second trial E_2 will be regarded as successful, if the sum of the counters extracted at the second, third and fourth drawings is greater than five, etc. In every case the result of the n trials will be expressed by n numbers, each of them equal to 0 or 1. The result $(1, 1, 0, 0, 0, \cdots 1)$, for instance, means that the first, the second, and the last trial were successful, the third, fourth, $\cdots$ unsuccessful, and we have an arithmetical probability distribution $v(x_1, x_2, \cdots x_n)$ $(x_k = 0, 1;$ $k = 1, 2, \cdots n)$, where

$$\sum_{x_1} \cdots \sum_{x_n} v(x_1, x_2, \cdots x_n) = 1. \tag{1}$$

Instead of the $2^n - 1$ values of v we will deal with certain groups of *partial sums* of them; the first is

$$\sum \cdots \sum v(x_1, x_2, \cdots x_{i-1}, 1, x_{i+1} \cdots x_n) = p_i \qquad (i = 1, 2, \cdots n)$$

where p_i is the probability that the i-th trial will be successful. In an analogous manner let p_{ij} be the probability that the i-th and the j-th trial are both sucessful, p_{ijk} the probability that the i-th, j-th and k-th trials are simultaneously successful. Let us provisionally denote by $\Sigma^{(i)}$ an $(n-1)$-tuple sum over all variables, except x_i, by $\Sigma^{(i,j)}$ an $(n-2)$-tuple sum over all variables except x_i and x_j etc. We shall then have:

$$\begin{aligned} p_i &= \sum{}^{(i)} v(x_1, \cdots x_{i-1}, 1, x_{i+1}, \cdots x_n) \\ p_{ij} &= \sum{}^{(i,j)} v(x_1, \cdots x_{i-1}, 1, x_{i+1}, \cdots x_{j-1}, 1, x_{j+1}, \cdots x_n) \\ &\cdots\cdots\cdots\cdots \\ p_{12\cdots n} &= v(1, 1, \cdots 1). \end{aligned} \tag{2}$$

In the following these probabilities $p_i, p_{ij}, p_{ijk} \cdots$ will be assumed as *directly given.* There are

$$\binom{n}{1} + \binom{n}{2} + \cdots \binom{n}{n-1} + \binom{n}{n} = 2^n - 1$$

values of this kind and it is easily seen, that the partial sums (2) are *linearly independent.*

If, especially, the probability $v(x_1, x_2, \cdots x_n)$ *depends only on the number* of zeros amongst $x_1, x_2, \cdots x_n$, i.e. if

$$v(1, 0, \cdots 0) = v(0, 1, 0, \cdots 0) = \cdots = v(0, 0, \cdots 1)$$
$$v(1, 1, \cdots 0) = v(1, 0, 1, \cdots 0) = \cdots = v(0, 0, \cdots 0, 1, 1)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

the value of p_i is independent of i, the value of p_{ij} independent of i and j, and so on:

$$p_1 = p_2 = \cdots = p_n$$
$$p_{12} = p_{23} = \cdots = p_{n-1,n}$$

In the particular case of *independent* events we have only to deal with n probabilities, namely $p_1, p_2, \cdots p_n$. We have indeed $p_{ij} = p_i p_j$; $p_{ijk} = p_i p_j p_k \cdots p_{12\cdots n} = p_1 p_2 \cdots p_n$.

In the case of *chains* however, we need only know $(2n - 1)$ values, namely $p_1, p_2, \cdots p_n$; $p_{12}, p_{23}, \cdots p_{n-1,n}$. The other p_{ij}, and the $p_{ijk}, \cdots p_{12\cdots n}$ can be expressed in terms of the above probabilities.

Returning now to the general case it is easily seen that in the expression for $P_n(x)$ the $p_i, p_{ij} \cdots$ will appear only in the following combinations

$$(3)\quad S_n(0) = 1, \qquad S_n(1) = \sum_{i}^{1\cdots n} p_i, \qquad S_n(2) = \sum_{i,j}^{1\cdots n} p_{ij}, \cdots S_n(n) = p_{12\cdots n}.$$

Indeed, at the basis of the solution of the "problem of sums," there are the following relations [11] between the $S_n(z)$ and the $P_n(x)$.

$$(4)\qquad S_n(z) = \sum_{x=z}^{n} \binom{x}{z} P_n(x) \qquad \begin{pmatrix} x = 0, \cdots n \\ z = 0, \cdots n \end{pmatrix}$$

The linear equations (4) may be solved (by recurrence) for the $P_n(x)$ and we find the important result that

$$(5)\qquad P_n(x) = \sum_{z=x}^{n} (-1)^{z+x} \binom{z}{x} S_n(z)$$

Let $M_n^{(z)}$ be the z-th *factorial moment* of $P_n(x)$, i.e.

$$(6)\qquad M_n^{(z)} = \sum_{x=z}^{n} x(x-1) \cdots (x - z + 1) P_n(x).$$

Making use of (4) and (6) we obtain

$$(7)\qquad M_n^{(z)} = z! S_n(z).$$

Our aim is to obtain information concerning the asymptotic behaviour of $P_n(x)$ by studying that of the moments of $P_n(x)$. The moments however are easily seen to be given in terms of the $S_n(z)$.

3. The asymptotic behavior of $P_n(x)$. Convergence towards the normal distribution.

a. The Principal Theorem. According as the mean value

$$M_n^{(1)} = S_n(1) = a_n = \sum_{x=1}^{n} xP_n(x) \tag{8}$$

remains bounded or not for indefinitely increasing n, there are two types of passage to a limit. In the first case the distribution will converge (under certain conditions) towards a *Poisson* distribution; in the second case it will approach (under certain conditions) a normal distribution. As regards the convergence towards the Poisson distribution the author has published [9] a sufficient condition which seems to be quite simple and general. We shall, however, not resume this problem in the present paper.

We propose, indeed, to prove in the following pages a new theorem concerning the convergence of

$$V_n(x) = \sum_{t \leq x} P_n(t)$$

towards a *normal distribution.*

For this purpose we introduce the following function of the discontinuous variable $z = 0, 1, 2, \cdots n$

$$g_n(z) = \frac{z+1}{a_n} \frac{S_n(z+1)}{S_n(z)} \tag{9}$$

or, more concisely written $g_z = \dfrac{z+1}{a} \dfrac{S_{z+1}}{S_z}$, where $S_n(z)$ is defined by (3). Putting $z = a_n u$, let us consider

$$g_n(a_n u) = h_n(u) \tag{10}$$

where u is regarded as a *continuous* variable in the interval from 0 to ϵ. ($\epsilon > 0$.)

Denoting the variance of x for $V_n(x)$ by $M_2 = s_n^2$ we shall prove the

THEOREM: *Let the function $h_n(u)$, defined by* (10) *satisfy the following conditions:*

(i) *If n is sufficiently large, $h_n(u)$ admits derivatives of every order in the interval* $(0, \epsilon)$

(ii) *At $u = 0$, the first derivative of $h_n(u)$ has a limit, for $n \to \infty$, which is different from* -1.

(iii) *If u is in the interval $(0, \epsilon)$ the k-th derivative of $h_n(u)$ remains, for every k, inferior to a bound N_k which is independent of n.*

Then

$$\lim_{n\to\infty} V_n'(a_n + ys_n\sqrt{2}) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{y} e^{-x^2} dx \tag{11}$$

We shall see that in many applications these conditions may reasonably be assumed as satisfied.

b. DEMONSTRATION OF THE THEOREM.

In order to prove the principal theorem, stated above, we shall at first deduce some properties of the (finite) *differences* of $g_n(z)$ $(z = 0, 1, \cdots)$ from the assumptions (i), (ii), (iii) which deal with the *derivatives* of $h_n(u)$. Indeed, the κ-th

difference of $g_n(z)$ with respect to z, (which contains the values of $g_n(z)$ for $z = 0, 1, \cdots \kappa$), differs only by the factor a_n^κ from the κ-th divided difference of $h_n(u)$, with respect to u (which is formed by the values of $h_n(u)$ for $u = 0, \frac{1}{a_n}, \cdots \frac{\kappa}{a_n}$. Let $n > \kappa$ and so large that $\kappa/a_n < \epsilon$; then all u-values used in the formation of the κ-th divided difference of $h_n(u)$ will be in the interval $(0, \epsilon)$. Now, as it is well known, the absolute value of any divided difference of order κ can not be larger than the largest derivative in an interval which contains all the abscissae, used in the formation of the divided difference. But according to hypothesis (iii) the κ-th derivatives of $h_n(u)$ in $(0, \epsilon)$ are all inferior to N_κ. Therefore[3] we have

$$|a_n^\kappa \Delta^\kappa g_n(z)| < N_\kappa \tag{12}$$

and for every $\gamma > 0$

$$\lim_{n\to\infty} a_n^{\kappa-\gamma} \Delta_{z=0}^\kappa g_n(z) = 0. \tag{13}$$

On the other hand from condition (ii) it follows, as is easily seen, that

$$\lim_{n\to\infty} a_n \Delta_{z=0} g_n(z) = a_n[g_n(1) - g_n(0)] = c \neq -1. \tag{14}$$

The equations (13) and (14) imply but *finite differences of* $g_n(z)$.

Let us now introduce certain new moments F_ν which we could call "factorial moments about the mean." They are indeed related to the factorial moments $M^{(\nu)}$ in exactly the same way as the moments M_ν about the mean are related to the moments M_ν^0 about the origin. Writing, S_z, a and g_z instead of $S_n(z)$, a_n and $g_n(z)$, we set

$$\begin{aligned} F_\nu &= \Delta_{z=0}^\nu (M^{(\nu)} a^{\nu-\nu}) = M^{(\nu)} - \nu M^{(\nu-1)} a + \binom{\nu}{2} M^{(\nu-2)} a^2 - \cdots \pm a^\nu \\ &= \nu! S_\nu - \nu! S_{\nu-1} a + \binom{\nu}{2} (\nu-2)! S_{\nu-2} a^2 - \cdots \pm a^\nu \end{aligned} \tag{15}$$

where, particularly,

$$F_0 = 1, \quad F_1 = 0. \tag{16}$$

From (15) we have:

$$\begin{aligned} M^{(\nu)} &= \nu! S_\nu = \sum_{s=0}^{\nu} F_{\nu-s} \binom{\nu}{s} a^s \\ &= F_\nu + \nu F_{\nu-1} a + \binom{\nu}{2} F_{\nu-2} a^2 + \cdots + \binom{\nu}{\nu-2} F_2 a^{\nu-2} + a^\nu \end{aligned} \tag{17}$$

Let us begin by proving the following

[3] If we only want to deduce (13) it is sufficient to suppose that N_k (without being independent of n) increases more slowly than any power of a_n.

LEMMA I: *It follows from* (13) *and* (14) *that we have for the* F_ν *defined by* (15)

$$\lim_{n\to\infty}\frac{F_\nu}{a^{\frac{1}{2}\nu}} = G_\nu = \begin{cases}0 & \text{if } \nu \text{ odd}\\ 1\cdot 3\cdots(\nu-1)c^{\frac{1}{2}\nu} & \text{if } \nu \text{ even.}\end{cases} \tag{18}$$

First we conclude from (15) and (14) that (18) is true for $\nu = 1$ and $\nu = 2$. In order to prove (18) for *every* ν, we shall point out, that

$$\lim_{n\to\infty}\frac{F_\nu}{a^{\frac{1}{2}\nu}} = (\nu-1)c\lim_{n\to\infty}\frac{F_{\nu-2}}{a^{\frac{1}{2}(\nu-2)}}\cdots \qquad (\nu = 2, 3, \cdots) \tag{19}$$

Setting

$$f_s = g_s - 1 \quad\text{and}\quad m_s = \frac{S_s z!}{a^s} = \frac{M^{(s)}}{a^s} \tag{20}$$

we get

$$g_s = \frac{m_{s+1}}{m_s} \tag{21}$$

and

$$\begin{aligned}\Delta m_s &= m_s f_s\\ \Delta^\nu m_s &= \Delta^{\nu-1}(m_s f_s)\end{aligned} \qquad (s = 0, 1, 2, \cdots) \tag{22}$$

But according to (15) we have

$$\underset{s=0}{\Delta^\nu} m_s = \frac{1}{a^\nu}F_\nu \tag{23}$$

and therefore

$$\frac{F_\nu}{a^{\frac{1}{2}\nu}} = a^{\frac{1}{2}\nu}\underset{s=0}{\Delta^{\nu-1}}(m_s f_s) = a^{\frac{1}{2}\nu}\sum f_{\alpha\beta}\Delta^\alpha m_s \Delta^\beta f_s = \sum f_{\alpha\beta}\frac{F_\alpha}{a^{\frac{1}{2}\alpha}}a^{\frac{1}{2}(\nu-\alpha)}\Delta^\beta f_s \tag{24}$$

$$(\alpha+\beta \geqq \nu-1;\ \alpha \leqq \nu-1,\ \beta \leqq \nu-1).$$

Here we have made use of the fact that the κ-th difference of a product uv can be transformed in a finite sum $\Sigma S_{\alpha\beta}\Delta^\alpha u\Delta^\beta v$ where α and β are non-negative integers and $\alpha \leqq \kappa$, $\beta \leqq \kappa$. (If we concern ourselves with derivatives and not with finite differences, we have, $\alpha + \beta = \kappa$ and $S_{\alpha\beta} = \binom{\kappa}{\alpha}$). Suppose

$$\alpha+\beta > \nu - 1.$$

Then $\beta \geqq \nu - \alpha$; therefore, as $\nu > \alpha$ we have $\beta > \frac{\nu-\alpha}{2}$. Since $\Delta^\beta f_s = \Delta^\beta g_s$ the product $a^{\frac{1}{2}(\nu-\alpha)}\underset{s=0}{\Delta^\beta} f_s$ *converges toward* zero, in accordance with (13), whereas the factor $S_{\alpha\beta}\frac{F_\alpha}{a^{\frac{1}{2}\alpha}}$ remains bounded for every $\alpha < \nu$. Now suppose

$$\alpha + \beta = \nu - 1.$$

Then $\beta = \nu - 1 - \alpha$. First let $\alpha < \nu - 2$; then $\beta = \nu - 1 - \alpha > \frac{\nu - \alpha}{2}$. Thus $a^{\frac{1}{2}(\nu-\alpha)}\Delta^{\beta}_{z=0} f_z$ converges again towards zero, whereas the other factors are bounded as before. Next, if $\alpha = \nu - 1$, then $\beta = 0$ and $\Delta^0_{z=0} f_z = f_0 = 0$. Thus the corresponding term of our sum is equal to zero. Finally if $\alpha = \nu - 2$, then $\beta = 1$, and $S_{\alpha\beta} = \nu - 1$. The corresponding term of the sum (24) will be

$$(\nu - 1) \lim_{n\to\infty} \frac{F_{\nu-2}}{a^{\frac{1}{2}(\nu-2)}} \cdot \lim_{n\to\infty} a \Delta_{z=0} f_z = (\nu - 1)c \lim_{n\to\infty} \frac{F_{\nu-2}}{a^{\frac{1}{2}(\nu-2)}}$$

which completes the proof of Lemma I.

We shall now establish a relation between the *factorial moments* about the mean F_ν and the *ordinary moments* about the mean M_ν. To an expression of the form

$$ca^{\rho}F_{\alpha} \tag{25}$$

(where the constant c is independent of n) let us attribute a "weight" $\rho + \frac{\alpha}{2}$. Then we shall prove the following *lemma*

LEMMA II: *Let* $\nu = 2\mu$ (ν even), $\nu = 2\mu + 1$ (ν odd) *and*

$$\alpha_\rho = \frac{\nu!}{(\nu - 2\rho)!\, 2^{\rho}\rho!}. \tag{26}$$

Then

$$M_\nu - \sum_{\rho=0}^{\mu} \alpha_\rho a^{\rho} F_{\nu-2\rho} \tag{27}$$

is equal to a finite sum of terms of the form (25), *each of which has a weight less than* $\nu/2$.

To prove this lemma we begin by expressing the M_ν in terms of the factorial moments $M^{(\rho)}$. We shall then express the $M^{(\rho)}$ by the F_z. Now, let $s_{\kappa z}$ be the "*Stirling* numbers of second kind," i.e., putting

$$x^{(z)} = x(x - 1) \cdots (x - z + 1) \tag{28}$$

we have

$$x^{z} = \sum_{\kappa=0}^{z} s_{\kappa z} x^{(z-\kappa)} \qquad (z = 0, 1, 2, \cdots) \tag{29}$$

Then by an elementary calculation we obtain

$$M_\nu = \sum_{\rho=0}^{\nu} M^{(\nu-\rho)} \left[s_{\rho\nu} - \nu a s_{\rho-1,\nu-1} + \binom{\nu}{2} a^2 s_{\rho-2,\nu-2} - \cdots \pm \binom{\nu}{\rho} a^{\rho} \right]. \tag{30}$$

If we now introduce the F_κ we get

$$M_\nu = \sum_{\rho=0}^{\nu-1} \sum_{r=0}^{\nu-\rho} F_{\nu-r-\rho} a^r \Big[\binom{\nu-\rho}{r} s_{\rho\nu} - \binom{\nu}{1}\binom{\nu-\rho-1}{r-1} s_{\rho,\nu-1} \tag{31}$$
$$+ \binom{\nu}{2}\binom{\nu-\rho-2}{r-2} s_{\rho,\nu-2} - \cdots \pm \binom{\nu}{r} s_{\rho,\nu-r} \Big].$$

Furthermore we may easily verify that

$$\binom{\nu-\rho-x}{r-x} = \binom{\nu-\rho}{r} - \binom{\nu-\rho-1}{r} x \tag{32}$$
$$+ \frac{1}{2!}\binom{\nu-\rho-2}{r} x^{(2)} + \cdots \pm \frac{1}{(\nu-r-\rho)!} x^{(\nu-r-\rho)}.$$

But the $s_{z z}$ for $z = 0, 1, 2, \cdots$ are equal to the values of a polynomial in z, of degree 2κ, the highest term of which is equal to $\frac{z^{2\kappa}}{\kappa!\,2^\kappa}$. The degree of the product

$$\binom{\nu-\rho-x}{r-x} s_{\rho,\nu-x} = \varphi(x) \tag{33}$$

is therefore equal to $(\nu - r - \rho) + 2\rho = \nu - r + \rho$. On the other hand the expression between brackets in the right hand member of (31) is nothing other than the ν-th difference of $\zeta(x)$. (The missing terms of this difference are indeed equal to zero, the corresponding $s_{z\nu}$ being equal to zero.)

This ν-th difference will certainly vanish if

$$\nu - r + \rho < \nu \text{ i.e. } r > \rho.$$

Now, let $r = \rho$. Then the ν-th difference, i.e. the coefficient α_ρ of $F_{\nu-r-\rho}a^r = F_{\nu-2\rho}a^\rho$ in (31), is equal to $\nu!$ multiplied by the coefficient of $x^{\nu-2\rho}$ in $\varphi(x)$:

$$\alpha_\rho = \nu! \frac{1}{(\nu-2\rho)!} \frac{1}{2^\rho \rho!}.$$

Finally, let $r < \rho$. Then the weight of $F_{\nu-r-\rho}a^r$ is inferior to $\nu/2$. We have thus established Lemma II.

We have for instance for $\nu = 1, 2, 3, 4, 5$

$M_1 = F_1 = 0,\ M_2 = F_2 + a,\ M_3 = F_3 + 3F_2 + a$

$M_4 = (F_4 + 6\, aF_2 + 3\, a^2) + 6\, F_3 + (7F_2 + a)$

$M_5 = (F_5 + 10\, F_3 a) + (10\, F_4 + 40\, F_2 a + 10\, a^2) + 25\, F_3 + (15\, F_2 + a)$

Inversely in an analogous manner, we can express F_ν by the

$$M_\rho (\rho = 1, 2, \cdots \nu).$$

We can now terminate our demonstration by proving the following

LEMMA III: *If the conditions* (18) *are satisfied, then*

$$\lim_{n\to\infty} \frac{M_\nu}{M_2^{\frac{1}{2}\nu}} = H_\nu = \begin{cases} 0 \cdots \nu \text{ odd} \\ 1\cdot 3 \cdots (\nu - 1) \cdots \nu \text{ even}. \end{cases} \tag{34}$$

First the equation (18) for $\nu = 2$ gives

$$\lim_{n\to\infty} \frac{F_2}{a} = \lim_{n\to\infty} \frac{M_2 - a}{a} = c$$

thus

$$\lim_{n\to\infty} \frac{M_2}{a} = 1 + c \qquad (c \neq -1). \tag{35}$$

It is therefore obviously sufficient to prove the relation

$$\lim_{n\to\infty} \frac{M_\nu}{a^{\frac{1}{2}\nu}} = H_\nu(1 + c)^{\frac{1}{2}\nu}. \tag{36}$$

Putting $\nu = 2\mu$ and $\nu = 2\mu + 1$ respectively we obtain however from our lemma

$$\frac{M_\nu}{a^{\frac{1}{2}\nu}} = \sum_{\rho=0}^{\mu} \alpha_\rho a^{\rho - \frac{1}{2}\nu} F_{\nu-2\rho} + R a^{-\frac{1}{2}\nu}.$$

Here R represents a finite sum of terms of the form (25), of "weight" inferior to $\frac{\nu}{2}$. But by virtue of (18) such a term, divided by $a^{\frac{1}{2}\nu}$ converges towards zero and we obtain

$$\lim_{n\to\infty} \frac{M_\nu}{a^{\frac{1}{2}\nu}} = \sum_{\rho=0}^{\mu} \alpha_\rho \lim_{n\to\infty} \frac{F_{\nu-2\rho}}{a^{\frac{1}{2}\nu-\rho}} = \sum_{\rho=0}^{\mu} \frac{\nu!}{(\nu - 2\rho)!\,2^\rho \rho!} G_{\nu-2\rho}. \tag{37}$$

For an odd ν, $G_{\nu-2\rho}$ *is equal to zero*; for an even $\nu (= 2\mu$, say) however, we have

$$G_{2\mu-2\rho} = c^{\mu-\rho} \frac{(2\mu - 2\rho)!}{2^{\mu-\rho}(\mu - \rho)!}$$

and we obtain

$$\begin{aligned} \lim_{n\to\infty} \frac{M_{2\mu}}{a^\mu} &= \sum_{\rho=0}^{\mu} \frac{(2\mu)!}{(2\mu - 2\rho)!\,2^\rho \rho!} \cdot \frac{(2\mu - 2\rho)!}{2^{\mu-\rho}(\mu - \rho)!} \\ &= \frac{(2\mu)!}{2^\mu \mu!} \sum_{\rho=0}^{\mu} \frac{\mu!}{\rho!(\mu - \rho)!} c^{\mu-\rho} = H_{2\mu}(1 + c)^\mu \end{aligned} \tag{38}$$

in accordance with (36). Lemma III is therefore proved.

Our principal theorem is now an obvious consequence of the well known theorem of the continuity of moments. By virtue of this theorem the convergence of $V_n(a_n + y s_n \sqrt{2})$ towards a normal distribution as given by (7) will indeed be assured if the moments of V_n converge towards the moments of the corresponding normal distribution; i.e. if (34) is true. Thus our principal theorem is completely demonstrated.

4. Some applications.

EXAMPLE 1. We shall consider the following play as a very simple application of our theorem: An urn contains $m = 2n$ counters bearing the numbers $1, 2, \cdots m$. We draw them all, one after the other, without returning the counters previously drawn. *We ask for the probability $P_{2n}(x)$ that an even counter will appear at a drawing of even number x times* $(0 \leqq x \leqq n)$.
As can be easily found, we have

$$p_2 = p_4 = \cdots = p_{2n} = \tfrac{1}{2}$$

$$p_{2,4} = p_{2,6} = \cdots = p_{2n-2,2n} = \frac{1}{4}\frac{2n-2}{2n-1}$$

Consequently

$$(39)\quad \begin{aligned} &S_1 = \frac{n}{2}, \qquad S_2 = \binom{n}{2}\frac{1}{4}\frac{2n-2}{2n-1}, \qquad S_3 = \binom{n}{3}\frac{1}{8}\frac{(2n-2)(2n-4)}{(2n-1)(2n-2)}, \\ &\cdots\cdot \\ &S_z = \frac{1}{2^z}\binom{n}{z}\frac{(2n-2)(2n-4)\cdots(2n-2z+2)}{(2n-1)(2n-2)\cdots(2n-z+1)}. \end{aligned}$$

From (39) it follows that

$$(40)\qquad g_n(z) = \frac{n-z}{n}\frac{2n-2z}{2n-z}.$$

Setting $z/\frac{1}{2}n = u$, we get

$$(41)\qquad h_n(u) = \frac{(2-u)^2}{2\left(2-\dfrac{u}{2}\right)}.$$

The conditions (i), (ii), (iii) of our principal theorem are obviously satisfied if $\epsilon < 4$ and we have

$$h_n'(0) = -\tfrac{3}{4} = c$$

The probability defined above is thus seen to converge (according to (11)) *towards a normal distribution, having a mean equal to $\dfrac{n}{2}$ and a variance $M_2 \sim \dfrac{n}{8}$*

EXAMPLE 2. *Probability of an "occupation."* Let k stones be distributed by chance over n places. Then the probability that any stone will occupy a certain place will be equal to $1/n$. *We ask for the probability $P_n(x)$ that there shall be x places, every one of which is occupied by exactly m stones.*[4]

By certain simple considerations, well known in combinatory calculus, we obtain:

$$(42)\qquad a_n = n\frac{k!}{m!(k-m)!}\left(\frac{1}{n}\right)^m\left(1-\frac{1}{n}\right)^{k-m}$$

---

[4] The problem presents itself for instance if we ask for the probability that in a certain county there will be x villages, everyone of m inhabitants.

$$S_z = \frac{n!}{z!(n-z)!}\,\frac{k!}{(m!)^z(k-mz)!}\left(\frac{1}{n}\right)^{mz}\left(1-\frac{z}{n}\right)^{n-mz}. \tag{43}$$

Let $k/n = \alpha$. From (43) we deduce that

$$\begin{aligned} g_n(z) = \frac{n-z}{n}&\left(\frac{1-\dfrac{z+1}{n}}{1-\dfrac{z}{n}}\right)^{n\alpha}\frac{1}{\left(1-\dfrac{1}{n}\right)^{n\alpha}}\,\frac{\left(1-\dfrac{z}{n}\right)^{mz}\left(1-\dfrac{1}{n}\right)^{m}}{\left(1-\dfrac{z+1}{n}\right)^{mz+m}} \\ &\cdot\frac{\left(\alpha-\dfrac{zm}{n}\right)\left(\alpha-\dfrac{zm+1}{n}\right)\cdots\left(\alpha-\dfrac{zm+m-1}{n}\right)}{\alpha\left(\alpha-\dfrac{1}{n}\right)\cdots\left(\alpha-\dfrac{m-1}{n}\right)} \end{aligned} \tag{44}$$

Now, let n *and* k *tend simultaneously to* ∞, *in such a way that* $\alpha = \dfrac{k}{n}$ *remains bounded*. We get at first

$$\lim_{n\to\infty}\frac{a_n}{n} = \frac{\alpha^m}{m!}e^{-\alpha}. \tag{45}$$

As a_n is seen to be of the order of magnitude of n we introduce the new variables

$$\frac{z}{n} = v \quad\text{and}\quad v = u\,\frac{a_n}{n}.$$

We have then (writing h and $\bar{h}$ instead of h_n and $\bar{h}_n$):

$$g_n(z) = g_n(nv) = \bar{h}(v)$$

$$\bar{h}(v) = \bar{h}\left(u\,\frac{a_n}{n}\right) = h(u).$$

Therefore

$$\begin{aligned} \bar{h}(v) = (1-v)&\left(1-\frac{1}{n(1-v)}\right)^{n\alpha}\frac{1}{\left(1-\dfrac{1}{n}\right)^{n\alpha}}\left(\frac{1-\dfrac{1}{n}}{1-v-\dfrac{1}{n}}\right)^{m} \\ \left(1-\frac{1}{n(1-v)}\right)^{-nmv}&\cdot\frac{(\alpha-mv)\left(\alpha-\dfrac{1}{n}-mv\right)\cdots\left(\alpha-\dfrac{m-1}{n}-mv\right)}{\alpha\left(\alpha-\dfrac{1}{n}\right)\cdots\left(\alpha-\dfrac{m-1}{n}\right)}. \end{aligned} \tag{46}$$

These formulae show that the k-th derivative of $\bar{h}(v)$ with respect to v contains only rational expressions, [in the denominators of which there appear powers of $(1-v)$], and positive powers of $\log\left(1-\dfrac{1}{n(1-v)}\right)$. The conditions (i) and (iii) of our principal theorem are therefore satisfied if $\epsilon < 1$. Furthermore we have

$$\left(\frac{dh}{dv}\right)_{v=0} = -1 - \frac{\alpha}{1-\frac{1}{n}} - mn\log\left(1-\frac{1}{n}\right) + \frac{m}{1-\frac{1}{n}}$$

(47)
$$-m\,\frac{\left(\alpha-\frac{1}{n}\right)\left(\alpha-\frac{2}{n}\right)\cdots\left(\alpha-\frac{m-1}{n}\right)+\alpha\left(\alpha-\frac{2}{n}\right)\cdots\left(\alpha-\frac{m-1}{n}\right)+\cdots\alpha\left(\alpha-\frac{1}{n}\right)\cdots\left(\alpha-\frac{m-2}{n}\right)}{\alpha\left(\alpha-\frac{1}{n}\right)\cdots\left(\alpha-\frac{m-1}{n}\right)}$$

and consequently

$$\lim_{n\to\infty}\left(\frac{dh}{du}\right)_{u=0} = \left[-1-\frac{m^2}{\alpha}-\alpha+2m\right]\lim_{n\to\infty}\frac{a_n}{n}$$
$$= -\left(1+\frac{(m-\alpha)^2}{\alpha}\right)\frac{\alpha^m}{m!}e^{-\alpha} = c.$$

We have thus obtained the interesting result that,
The probability $V_n(x)$ that x places at most are occupied, each one by m stones, converges towards a normal distribution if k and n tend simultaneously to ∞ in such a way that $\lim_{n\to\infty}\frac{k}{n} = \alpha$ is bounded. We have then

(48)
$$\lim_{n\to\infty} V_n(a_n + u\sqrt{2}\, s_n) = \phi(u)$$

with

(49)
$$\lim_{n\to\infty}\frac{a_n}{n} = \frac{\alpha^m}{m!}e^{-\alpha}, \qquad \lim_{n\to\infty}\frac{s_n^2}{a_n} = 1 - \frac{\alpha^m e^{-\alpha}}{m!}\cdot\left[1+\frac{(m-\alpha)^2}{\alpha}\right].$$

UNIVERSITY OF ISTANBUL,
ISTANBUL, TURKEY.

REFERENCES

[1] H. POINCARÉ, *Calcul des Probabilités*, Paris, 1912
[2] W. BURNSIDE, *Theory of Probability*, Cambridge, 1928
[3] G. U. YULE, *Introduction to the Theory of Statistics*, London, 1932
[4] C. JORDAN, *Acta Litter, Scient.* (Szégéd), Vol. III (1927), p. 193
[5] C. JORDAN, *Acta Litter. Scient.* (Szégéd), Vol. VII (1934), p. 103
[6] E. J. GUMBEL, *Comptes Rendus*, Vol. 202, (1936), p. 1627
[7] E. J. GUMBEL, *Giorn. Inst. Ital. Att.*, Vol. XVI (1938)
[8] H. GEIRINGER, *Comptes Rendus*, Vol. 204 (1937), p. 1856
[9] H. GEIRINGER, *Comptes Rendus*, Vol. 204 (1937), p. 1914
[10] H. GEIRINGER, *Revue Interbalconique*, (Athens), Vol. II, pp. 1–26
[11] R. VON MISES, *Zs. Aug. Math. und Mech.*, Vol. I (1921)

FIDUCIAL DISTRIBUTIONS IN FIDUCIAL INFERENCE*

By S. S. Wilks

1. Introduction. The essential idea involved in the method of argument now known as fiducial argument, at least in a very special case, seems to have been introduced into statistical literature by E. B. Wilson [1] in connection with the problem of inferring, from an observed relative frequency in a large sample, the true proportion or probability p associated with a given attribute. Since 1930 the ideas and terminology surrounding the fiducial method have been developed by R. A. Fisher [2, 3], J. Neyman [4, 5] and others into a system for making inferences from a sample of observations about the values of parameters which characterize the distribution of the hypothetical population from which the sample is assumed to have been drawn. The functional form of the population distribution law is assumed to be known The parameters may be means, a difference between means, variances, ranges, regression coefficients, probabilities or any other descriptive indices or combinations of indices which may be considered important in specifying the distribution function of a population. In arguing fiducially about the value of a parameter, a procedure applicable to some of the simple cases begins by the calculation from the sample of an *estimate* of the parameter in question. The values of the estimate in repeated samples of the same size will theoretically cluster "near" the true value of the parameter according to a certain distribution law which can, in general, be deduced from the functional form of the population distribution law. If the distribution of the estimate involves only the one parameter, and if, as is frequently the case, one can find a function ψ of the estimate and the parameter which has a distribution not depending on the parameter, then one is able to set up, in a rather simple manner, *fiducial limits* or a *confidence interval* for the parameter corresponding to the observed value of the estimate. The limits will depend on the particular method of calculating the estimate, the value of the estimate in the sample, and on the degree of risk of being wrong which one is willing to take in stating that the limits will include between them the value of the parameter for the population under consideration. In general the smaller the degree of risk, the wider apart will be the limits. Thus for a given pair of limits there will be an associated degree of uncertainty that the true value of the parameter is actually included between those limits. This uncertainty can be expressed by a probability α calculated from the sampling distribution of the ψ function of the parameter and estimate. Under certain conditions, one can, by simply changing variables, obtain from the ψ

* An expository paper presented to the American Statistical Association on December 28, 1937, at the invitation of the Program Committee.

distribution what has been termed by Fisher a *fiducial distribution* function of the parameter. From the fiducial distribution and for a given value of the estimate one can actually determine fiducial limits of the parameter corresponding to a given risk α. It will be seen as we proceed that the fiducial distribution plays no indispensable part in fiducial inference; the ψ function and its distribution from which the fiducial distribution is derivable, are sufficient for the fiducial argument in many cases that commonly arise in statistics. We shall discuss fiducial argument and fiducial distributions from the point of view of ψ functions.

2. Example. To illustrate these points let us consider an example, namely, the problem of determining fiducial limits and the fiducial distribution of the range of a rectangular distribution for a given value of the range in a sample "randomly drawn" from it.

If a sample of n individuals is drawn from a population whose distribution law is $f(x, \theta) = 1/\theta$, where only values of x between 0 and θ are considered, (that is, a rectangular distribution having range θ) the probability that the range r of the sample lies between r and $r + dr$ is $\varphi(r, \theta)\, dr$, where

$$\varphi(r, \theta) = \frac{n(n-1)}{\theta^n}(\theta - r)r^{n-2}. \tag{1}$$

Here θ is the parameter under question, and r is the estimate; r is the difference between the largest and smallest variate in the sample. Thus, for a given value of θ, say θ_0, $\varphi(r, \theta_0)$ is a sampling distribution law defined for given values of r on the range $r = 0$, to $r = \theta_0$. If we let $r/\theta = \psi$, then

$$\varphi(r, \theta)\, dr = n(n-1)(1-\psi)\psi^{n-2}\, d\psi = G(\psi)\, d\psi, \tag{2}$$

which, from a statistical point of view, shows that if we should take an aggregate of randomly drawn samples (of n items each) from rectangular populations and calculate ψ for each *sample-population combination*, then the distribution of ψ will be that given in (2). By a *sample-population combination* in this example we mean any rectangular population that may arise and a "randomly drawn" sample from it. The possible values of ψ range from 0 to 1. Thus if ψ_α is such that

$$n(n-1)\int_0^{\psi_\alpha}(1-\psi)\psi^{n-2}\,d\psi = \alpha, \quad \text{i.e.} \quad \psi_\alpha^{n-1}[n-(n-1)\psi_\alpha] = \alpha, \tag{3}$$

and if we draw a sample of n from a rectangular population, we can claim that the probability is $1 - \alpha$ that the ψ produced by this sample-population combination will satisfy the inequality

$$\psi_\alpha < \psi < 1. \tag{4}$$

It should be observed that there are many pairs of numbers, say ψ'_α and ψ''_α such that we can claim that $\psi'_\alpha < \psi < \psi''_\alpha$, with probability $1 - \alpha$ of being

correct in making the claim. ψ'_α and ψ''_α are ordinarily chosen so that the interval formed by them is as short as possible (or approximately so) in some sense. Inequality (4) is equivalent to each of the following inequalities

$$\psi_\alpha < \frac{r}{\theta} < 1, \qquad \frac{r}{\psi_\alpha} > \theta > r. \tag{5}$$

Now ψ_α can be determined from (3) when n and α are given. For example, if $\alpha = .01$ and $n = 10$, we find from (3) that $\psi_\alpha = .495$. For a given sample, the fiducial limits r/ψ_α and r can be calculated from ψ_α and the sample. It will be noticed that fiducial limits are nothing more nor less than random variables that fluctuate from sample to sample. The interval between r and r/ψ_α is called a *confidence interval* or *fiducial interval*; $1 - \alpha$ is known as the *confidence coefficient* [4] associated with the limits. Hence, in repeated samples of n from a rectangular population with range θ_0, $100(1 - \alpha)$ percent of the samples will produce fiducial limits r/ψ_α and r which include the fixed value θ_0 between them. This statement holds regardless of the value of θ_0. Hence in an aggregate of sample-population combinations, the aggregate of pairs of fiducial limits r/ψ_α and r will, in $100(1 - \alpha)$ percent of the combinations, include between them the true value of the range of the population. Furthermore, whether there is only one rectangular population for all sample-population combinations or many different rectangular populations, this statement remains true, thus showing that the method of fiducial limits for inferring the value of the parameter is independent of any *a priori* distribution of rectangular populations in an aggregate of sample-population combinations—the distribution being with respect to values of θ.

Let us look at the matter geometrically. Suppose we are drawing samples from a rectangular population with $\theta = \theta_0$. The r for each sample is represented by a dot along Or in Figure 1; corresponding to each dot there is confidence interval cutting across the V-shaped region MOR. The probability is $1 - \alpha$ that a confidence interval computed from a sample from the population having range θ_0 will cut the line $\theta_0 K$. The cutting of $\theta_0 K$ by a confidence interval is equivalent to the statement that θ_0 is included between the corresponding fiducial limits.

From a practical statistical point of view what we have said has the following meaning: If on each occasion in which a randomly drawn sample of n from some rectangular population is considered, one (i) calculates the numbers r/ψ_α and r, and (ii) asserts that the range in the population producing the sample lies between these two computed limits, then in about $100(1 - \alpha)$ percent of the cases assertion (ii) will be correct (theoretically). Thus, in dealing with samples of 10 individuals from rectangular populations, one would be correct (theoretically) in about 99 percent of the cases by asserting that the population range will lie between the sample range and $2.020 \left(= \frac{1}{.495}\right)$ times the sample range. More generally, one need not use the same value of n all the way

through, provided that for the given α one evaluates ψ_α according to (3), for each n that arises. It will be seen from (3) that as n increases, the value of ψ_α tends to 1 and hence the fiducial limits r/ψ_α and r for any given sample tend to the same value, namely the sample range, thus showing that fiducial inferences about θ can be made arbitrarily certain by taking sufficiently large samples.

It is evident that the method of fiducial limits furnishes a satisfactory procedure for inferring the value of the population range θ from samples drawn from rectangular populations. Let us now go a step further and consider the fiducial distribution of θ and how it fits into the scene. The cumulative distribution of ψ is

$$\psi^{n-1}[n - (n-1)\psi] \tag{6}$$

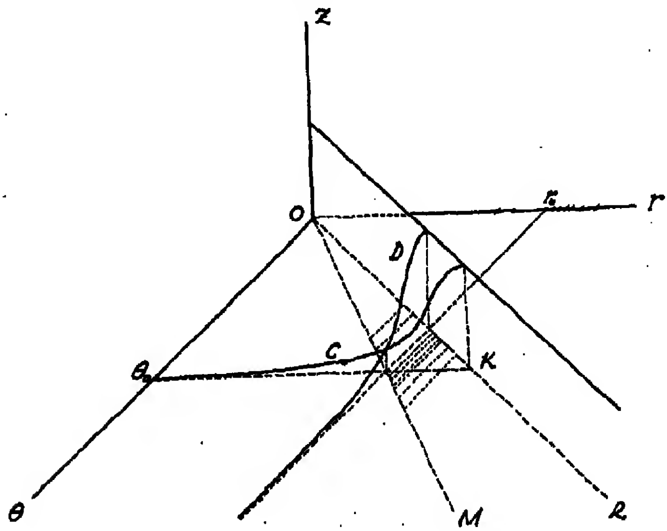

FIG. 1

and hence the cumulative distribution of r for a fixed θ, say θ_0, is

$$F(r, \theta_0) = \left(\frac{r}{\theta_0}\right)^{n-1}\left[n - (n-1)\left(\frac{r}{\theta_0}\right)\right] \tag{7}$$

which increases from 0 to 1 as r increases from 0 to θ_0. Geometrically, $z = F(r, \theta)$ can be represented as a surface defined over the region bounded by lines $O\theta$ and OR in Figure 1, such that z is zero along O and is unity along the line OR $(r = \theta)$. $F(r, \theta)$ is continuous inside the region θOR, and for any given value $r_0 \neq 0$ of r, $F(r, \theta)$ decreases from 1 to 0 as θ increases from r_0 to ∞. The curves having the equations

$$\begin{cases} z = F(r, \theta) \\ \theta = \theta_0 \end{cases} \quad \text{and} \quad \begin{cases} z = F(r, \theta) \\ r = r_0 \end{cases}$$

(where θ_0, r_0, and α are such that $r_0/\theta_0 = \psi_\alpha$ and $F(r_0, \theta_0) = \alpha$) are the curves C and D respectively. C is the cumulative distribution of ranges of samples of n from a rectangular population with range θ_0. The curve D has the mathematical characteristics of a cumulative distribution function cumulated in the negative direction with respect to θ: its ordinates increase from 0 to 1 as θ decreases from ∞ to θ_0. Thus, if we take $-\frac{\partial}{\partial\theta}F(r_0, \theta)$ we get a function $g(\theta, r_0)$ which has the essential mathematical characteristics of a distribution function: it is non-negative, can be integrated over any interval of θ, and has total area under it equal to unity. We have

$$(8) \qquad g(\theta, r_0) = n(n-1)\frac{r_0^{n-1}}{\theta^n}\left(1 - \frac{r_0}{\theta}\right)$$

and it is called the *fiducial distribution* of θ for $r = r_0$. It must be firmly pointed out that θ is not a random variable and hence $g(\theta, r_0)$ is *not* a distribution function of a random variable, although it has the mathematical properties of such a distribution. Objections have been raised to the use of the term fiducial distribution on the grounds that the thing to which it applies is not a distribution at all. However, as long as the term is carefully defined there should be no ambiguity in using it. From an analytical point of view, the problem of obtaining the fiducial distribution of θ is only a matter of changing variables for since

$$(9) \qquad \varphi(r, \theta)\, dr = g(\theta, r)\, d\theta = n(n-1)(1-\psi)\psi^{n-2}\, d\psi$$

and $\psi_\alpha = r_0/\psi_0$, we have

$$(10) \quad \int_{\theta_0\psi_\alpha}^{\theta_0} \varphi(r, \theta_0)\, dr = \int_{r_0}^{r_0/\psi_0} g(\theta, r_0)\, d\theta = \int_{\psi_\alpha}^{1} n(n-1)(1-\psi)\psi^{n-2} d\psi = 1 - \alpha.$$

We remark again that

$$(11) \qquad \int_{r_0}^{\theta_1} g(\theta, r_0)\, d\theta$$

is *not* to be interpreted as probability as though θ were a random variable. Instead, the meaning is as follows: Let r_0 be the range in a sample known to be from *some* rectangular population, and let the value of r_0 be inserted in (11), and let θ_1 be determined so that the value of the integral is $1 - \alpha$. The two limits for the integral are fiducial limits associated with the sample for the confidence coefficient $1 - \alpha$, which were discussed earlier. Thus, for each sample, we can compute fiducial limits using the fiducial distribution. These limits, as we have seen by considering the ψ function, fluctuate from sample to sample in such a way that the probability is $1 - \alpha$ that they will include between them the true value of the range of the population under consideration.

3. Summary of Principles. From the point of view we have taken the essential notions involved in the method of fiducial argument and fiducial

distributions for the case of a continuous variate and one parameter can be readily abstracted from the example just discussed. In general, we have the following steps:

(a) A sample is assumed to be *randomly drawn* from a population with a distribution of *known* functional form $f(x, \theta)$, θ being a parameter. Let $x_1, x_2, \cdots, x_n$ be the values of x in the sample.

(b) A function, say $\psi(x_1, \cdots, x_n, \theta)$ of the sample x's and θ is devised so that its sampling distribution $G(\psi)$ involves θ and the x's only as they enter into ψ. The value of θ in ψ is that for the population from which the sample is actually drawn.

(c) Two numerical values of ψ, say ψ'_α and ψ''_α are chosen (ordinarily as close together as possible) so that the probability computed from $G(\psi)$ is $1 - \alpha$ (e.g. 0.95) that ψ will lie between ψ'_α and ψ''_α—more briefly $P(\psi'_\alpha < \psi < \psi''_\alpha) = 1 - \alpha$.

(d) The inequality $\psi'_\alpha < \psi < \psi''_\alpha$ which contains only one unknown, namely θ, is solved for θ giving the equivalent inequality $\underline{\theta} < \theta < \bar{\theta}$ where θ and $\bar{\theta}$ are *fiducial limits* and are subject to sampling fluctuations.

(e) The expression $P(\psi'_\alpha < \psi < \psi''_\alpha) = 1 - \alpha$ is replaced by the equivalent expression $P(\underline{\theta} < \theta < \bar{\theta}) = 1 - \alpha$ which states that the probability is $1 - \alpha$ that a sample will yield values $\underline{\theta}$ and $\bar{\theta}$ which will include the true value of θ between them.

(f) The differential element for the fiducial distribution of θ is $G(\psi) \left| \dfrac{\partial\psi}{\partial\theta} \right| d\theta$ (provided $\partial\psi/\partial\theta$ is a function of θ which does not change sign for a given sample of x's) and is obtained by letting θ be the variable in $G(\psi)\, d\psi$, keeping the x's fixed.

To give precisely the conditions under which all of these steps can be performed is a technical matter which will not be considered here. It is sufficient to remark that they can be performed in many cases of practical interest. Fiducial argument can be carried on using only the first five steps without introducing the notion of a fiducial distribution. In connection with step (a) it should be particularly noticed that the functional form $f(x, \theta)$ of the population under question is assumed to be known and that the sample under consideration is "randomly drawn" from the population. Thus, in applying the theory to practical problems it is a matter of fundamental importance that these two assumptions be valid. In cases where a sufficient amount of data exists, it can usually be satisfactorily tested by using the χ^2 test and other devices, whether or not a given functional form for $f(x, \theta)$ is a valid assumption. In cases where sufficient data do not exist for actually making such a test justification for assuming a given function form usually has to be made on the basis of theoretical considerations. From a practical point of view the notion of randomness is characterized by methods of drawing samples rather than *a posteriori* mathematical considerations of the sample after it has been drawn, and thus the question of randomly drawing samples depends largely upon the

experience and sound judgment of the experimenter. However, after one or more samples have been drawn "at random," the problem of arguing from them about the populations from which they were drawn is largely mathematical.

4. Case of large samples. For a population with a distribution of known functional form, a fiducial distribution of the parameter clearly depends on the size of the sample and the particular estimate used. For example, in large samples, we would get a fiducial distribution of the mean of a normal population of known variance by using the sample mean which would be different from the one obtained using the median of the sample. In order to be able to make the inferences about θ as accurate as possible, a ψ function should theoretically be used which will produce fiducial limits which are closest together, on the average, or perhaps "best" in some other sense, for a given α. The fiducial distribution obtainable from such a ψ could then be referred to as the "best" fiducial distribution, and theoretically it should be used in preference to other possible fiducial distributions if fiducial distributions are to be used at all to set fiducial limits. In large samples from a population with a distribution function $f(x, \theta)$, it is known [6] that, under rather general conditions, fiducial limits which are closest together on the average can be obtained by letting

$$\psi = \frac{1}{\sqrt{n}}\left(\frac{\partial L}{\partial \theta}\right)\left[E\left\{\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right\}\right]^{-\frac{1}{2}} \tag{12}$$

and treating ψ as a normally distributed variate with zero mean and unit variance, where $L = \sum_{i=1}^{n} \log f(x_i, \theta)$, the logarithm of the likelihood of θ for the given sample, $x_1, x_2, \cdots, x_n$ are values of x in the sample, and E denotes mathematical expectation. For example, in the case of a binomial population where each individual belongs either to class A or class B, we have $f(x, \theta) = \theta^x(1 - \theta)^{1-x}$ where θ is the probability associated with class A, x will be 0 or 1 according to whether an individual belongs to B or A. In a sample of n individuals, $L = m \log \theta + (n - m) \log (1 - \theta)$, where m is the number of individuals in class A. $E\left\{\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right\} = \frac{1}{\theta(1 - \theta)}$, and we get $\psi = \frac{m - n\theta}{\sqrt{n\theta(1 - \theta)}}$. If we should want to find fiducial limits of θ for a confidence coefficient of .95, we would solve (1) the equations $\frac{m - n\theta}{\sqrt{n\theta(1 - \theta)}} = \pm 1.96$ for θ, thus getting two values of θ, say $\underline{\theta}$ and $\bar{\theta}$. We can then say that $\underline{\theta}$ and $\bar{\theta}$ will include the true value of θ between them with a probability of .95 of being correct, in the sense that if we applied this rule consistently to samples from binomial populations, we would have a procedure that would lead to a correct statement in about 95 percent of the cases (theoretically).

To illustrate the difference between the fiducial method and the commonly

used method of placing limits on θ for $P = .95$, consider an example in which $m = 150$, $n = 400$. The usual procedure is to replace θ by m/n in $\theta \pm 1.96 \sqrt{\frac{\theta(1-\theta)}{n}}$, which yields .311 and .431. The fiducial procedure is to solve the equation $\frac{m - n\theta}{\sqrt{n\theta(1-\theta)}} = \pm 1.96$, for θ, thus obtaining .312 and .455. For the case of small samples, the problem of getting "best" fiducial limits becomes more complicated [5].

5. **Extensions of Fiducial Argument.** It will be observed that it is not necessary for ψ to be a function of only one statistic and θ in order to be able to argue fiducially about θ. For example, if a sample of n is drawn from a normal population with mean θ, it is well known that if $\bar{x}$ is the sample mean then

$$\psi = \frac{(\bar{x} - \theta)\sqrt{n(n-1)}}{\left[\sum_1^n (x_i - \bar{x})^2\right]^{\frac{1}{2}}} \tag{13}$$

(which is Fisher's t function), has the distribution

$$\frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}(n-1))\sqrt{\pi(n-1)}} \frac{d\psi}{[1 + \psi^2/(n-1)]^{\frac{1}{2}n}}. \tag{14}$$

Here ψ is a function of two statistics, namely $\bar{x}$ and $\sum_{i=1}^n (x_i - \bar{x})^2$, and the fiducial distribution of θ for this ψ function is obtained at once by applying rule (f).

The ideas of fiducial argument may be extended in other directions, but these cannot be considered in any detail here. For example ψ may be a function of $x_1, \cdots, x_n$ and two or more population parameters, in which case one could set up fiducial *regions* for the several parameters. From a practical point of view, the fiducial argument for two or more parameters simultaneously, had hardly been touched. Again ψ may be a function of statistics from two samples, one observed and the other not yet observed, and not involving population parameters, at all, in which case one can argue fiducially about the statistic in question for the unobserved sample [3]. The notion of a fiducial distribution has been extended to several parameters taken simultaneously [3, 7], but the problem of working out relations between fiducial distributions of several parameters and fiducial regions is yet to be investigated. The principles may be readily applied in situations in which the x's involved in ψ take on discrete values. In this case the equality signs in the probability expressions in steps (c) and (d) would be replaced by greater than or equal signs ($\geq$). Two excellent examples of the application of principles of fiducial argument to the discrete case are furnished: (i) by a paper by Pearson and Clopper [8] on fiducial limits of the probability P from samples from a binomial population, and (ii) by a paper by Ricker [9] on fiducial limits of m in the Poisson distribution $f(x, m) = m^x e^{-m}/x!$.

REFERENCES

[1] E. B. WILSON, "Probable Inference, the Law of Succession, and Statistical Inference," *Jour. American Statistical Association*, Vol. 22 (1927), pp. 209-212.

[2] R. A. FISHER, "The Concepts of Inverse Probability and Fiducial Probability Referring to Unknown Parameters," *Proc. Royal Society of London*, Series A, Vol. 139 (1933), pp. 343-348.

[3] R. A. FISHER, "The Fiducial Argument in Statistical Inferences," *Annals of Eugenics*, vol. 6 (1935), pp. 391-398.

[4] J. NEYMAN, "On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection," *Royal Statistical Society*, vol. 97, 1934, pp. 558-625.

[5] J. NEYMAN, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," *Phil. Trans. Roy. Soc. London*, Series A, Vol. 236 (1937), pp. 333-380.

[6] S. S. WILKS, "Shortest Average Confidence Intervals from Large Samples," *Annals of Mathematical Statistics*, Vol. 9 (1938), pp. 166-175.

[7] I. E. SEGAL, "Fiducial Distribution of Several Parameters with Application to a Normal System," *Proc. Cambridge Phil. Soc.*, vol. 34 (1938), pp. 41-47.

[8] C. J. CLOPPER AND E. S. PEARSON, "The Use of Confidence or Fiducial Limits in the Case of the Binomial," *Biometrika*, vol. 24, 1934, pp. 404-413.

[9] WILLIAM E. RICKER, "Fiducial Limits of the Poisson Frequency Distribution," *Jour. American Statistical Association*, vol. 32, 1937, pp. 349-356.

PRINCETON UNIVERSITY.

BIOLOGICAL APPLICATIONS OF NORMAL RANGE AND ASSOCIATED SIGNIFICANCE TESTS IN IGNORANCE OF ORIGINAL DISTRIBUTION FORMS*

By William R. Thompson

The word *normal* has been used in many senses—commonly by statisticians to designate a well-known distribution function. Another use familiar to biologists, particularly in experimental work and medicine, is to denote an untreated or control part of a universe, or a part whose members are free from specified characteristics such as evidence of past or present disease or malformation. Closely related to this last usage are attempts to delimit so-called normal ranges of variation for a quantitative attribute of the members of part or all of a universe in question. Interpretations are often vague, as when the interval between the least and greatest values observed in either a large or a small number of instances is taken to estimate a normal range. We shall consider the problem of using ranked data for estimating normal ranges as defined in the next paragraph.

If the instances have been drawn at random from a universe (U) of all possible observations obtainable in a prescribed manner, and are enumerated in ascending order of magnitude, $\{x_i\}$ for $i = 1, \cdots, n$; then it is proposed to show in the present communication how ranges of the type (x_k, x_{n+1-k}) may be used to estimate *normal ranges*, R_f, where the subscript f is the theoretical probability that a random value, x, drawn from U will lie within the range R_f, g that it will lie above, and g that it will lie below (where $2g = 1 - f$). Furthermore, it is proposed to show how these ranges may be used as the basis of significance tests where altered conditions appear to lead to abnormal biological variation. The form of frequency-distribution of U is supposed unknown, and is without effect upon the analysis. *Section* 1 is a development of the theory of range estimation, treated briefly in a previous paper [1] together with illustrations of its application. *Section* 2 deals with significance tests.

1. **The Method of Range Estimation.** Let x be a real variate, a random value drawn from an infinite universe or population U. Let $f(x)$ be the frequency function of x in U, supposed unknown; and $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Then for any given α and β, where $\alpha < \beta$, and

$$(1) \qquad P(\alpha < x < \beta) \equiv \int_{\alpha}^{\beta} f(x)\,dx.$$

* Presented at a meeting of the American Statistical Association, December 28, 1937, Atlantic City, N. J.

To facilitate development, suppose that in any finite sampling under consideration no two values of x may be exactly the same. Let $S = \{x_k\}$, $k = 1, \cdots, n$, denote a random sample from U, where the order of enumeration is arbitrary, but temporarily taken as a random order (to fix the ideas, consider this the order obtained in drawing). Let p_k be defined by

$$(2) \qquad p_k = P(x < x_k) \equiv \int_{-\infty}^{x_k} f(x)\,dx \quad \text{from which} \quad dp_k = f(x_k)\,dx_k.$$

Then p_k is the probability that a random x from U shall be less than any number x_k. Then obviously if x_k is drawn at random from U, p_k is a random variable whose distribution is the unit rectangle; i.e., $P(p' < p_k < p'') = p'' - p'$. Furthermore, the joint probability that x_k will lie in the interval x_k, $x_k + dx_k$ and that exactly r values in the sample S will be less than x_k is, to within terms of order dp_k, $\binom{n-1}{r} \cdot p_k^r \cdot (1 - p_k)^{n-1-r}\,dp_k$.

Then, in repeated sampling as above, for the case where just r of the n random values $\{x_i\}$ are less than the k-th drawn, let $P_{n,r}(p' < p_k < p'')$ denote the probability that p_k lies in the interval (p', p''). Then

$$(3) \qquad P_{n,r}(p' < p_k < p'') = \frac{(r+s+1)!}{r!\cdot s!} \cdot \int_{p'}^{p''} p^r \cdot q^s \cdot dp,$$

where $s = n - 1 - r$, and $q = 1 - p$. Obviously, the expression on the right of (3) does not depend on k if this index is the order of draft or a random index, but only upon the condition that exactly r of the n random values from U be less than a value x_k drawn at random from the sample of n values. Accordingly, we obtain the same result if we enumerate the n values $\{x_i\}$ in ascending order of magnitude $(x_i < x_j$, if $i < j)$. Then $k = r + 1$, in the cases considered, and (3) may be written,

$$(4) \qquad P_n(p' < p_k < p'') = \frac{n!}{(k-1)!(n-k)!} \cdot \int_{p'}^{p''} p^{k-1} \cdot q^{n-k} \cdot dp,$$

for $0 \leqq p' \leqq p'' \leqq 1$. Obviously, the result is the same if we deal instead with the k-th value (x_k) of every random sample S drawn. In passing it may be noted that for $p' = 0$ and $p'' = p$ in (4) we have

$$(5) \qquad P_n(p_k < p) = I_p(k, n - k + 1),$$

which may be evaluated for k, $n - k + 1 \leqq 50$ by means of the *Tables of the Incomplete Beta-Function* [2].

Of course, $P_n(0 < p_k < 1) = 1$, and (4) gives $\bar{p}_k$, the mean value of p_k in repeated random sampling of n values from U, as

$$(6) \qquad \bar{p}_k = \frac{n!}{(k-1)!(n-k)!} \cdot \int_0^1 p^k \cdot q^{n-k} \cdot dp = \frac{k}{n+1}.$$

Similarly, the variance, $\sigma^2_{p_k}$, of p_k is given by

$$\sigma^2_{p_k} = E[(p_k - \bar{p}_k)^2] = \frac{k(n-k+1)}{(n+1)^2\cdot(n+2)}. \tag{7}$$

Now suppose that we want to find a range (α, β) such that, in random drafts from U, the theoretical relative frequency of drawing x less than α is g, and the same as that of drawing x greater than β. (α, β) may be called a *central confidence range* with a *confidence* $f = 1 - 2g$ that x drawn at random from U will lie within the range. For $g = k/(n + 1)$ we may take the range $R_f = (x_k, x_{n-k+1})$; and likewise with $g = 5\%$ we may estimate, or approximate by interpolation where $20k > n + 1 > 20(k - 1)$, a range R_f for normal biological variation of a specified character, and this may be called briefly the estimated 90% *central normal range.*

Of course the probability of drawing $x < \alpha$ is $\int_{-\infty}^{\alpha} f(x)\,dx$, and that of drawing $x > \beta$ is $\int_{\beta}^{\infty} f(x)\,dx$; and these probabilities are unknown, as the frequency function $f(x)$ is unknown; but with $\alpha = x_k$ and $\beta = x_{n-k+1}$ the theoretical relative frequency in each case is $k/(n + 1)$ regardless of the universe.

It has been shown [1] also that if the sample S were drawn at random from a finite ordered population of aggregate number N, denoted by U_N, and Np_k is the number of values in U_N that are less than the k-th member of the given random sample in ascending order of magnitude; then, for S a sample of n values as before, the mean value of p_k in repeated sampling is

$$\bar{p}_k = \frac{k}{n+1}\left(1 + \frac{1}{N}\right) - \frac{1}{N}, \quad \text{and}$$

$$\sigma^2_{p_k} = \frac{k(n-k+1)}{(n+1)^2\cdot(n+2)}\cdot\left(1 + \frac{1}{N}\right)\left(1 - \frac{n}{N}\right).$$

An example is furnished by an analysis of data reported by Wadsworth and Hyman [3] in a study of influences of antigenic treatment of horses upon their plasma concentration of esterified cholesterol, free cholesterol, and phospholipids. As in chart 1 for normal horses, a graph has been constructed for each horse studied, using time as abscissa and a logarithmic ordinate scale for observed values of plasma concentration of the constituents:

1. Esterified Cholesterol,
2. Free Cholesterol, and
3. Phospholipids *times one-tenth,*

the respective successive points for each being joined to form three polygon curves. As these are in all cases discrete and lie in the order of enumeration from top to bottom of the graph, no special label seemed needed; but estimated normal ranges for the central 90% of variation have been indicated in each

case by two horizontal lines between brackets at the right, numbered to correspond with the enumeration above. The ranges are based on observations on 62 plasma samples, each from a different presumably normal horse. The normal horses in the chart show about the same individual variations; but, of course, the ranges are not to be interpreted to indicate normal variation for an individual animal.

Chart 2 presents in like manner the data obtained for horses under immunization against tetanus and the streptococcus. The tetanus immunization treat-

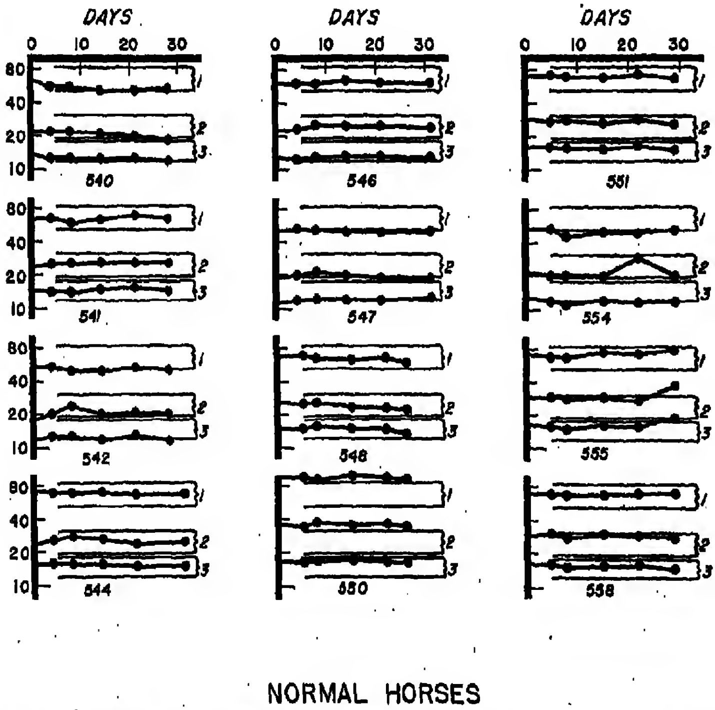

Chart 1. On each graph for a given normal horse, the number of which appears below, the curves in descending order respectively represent (1) esterified cholesterol, (2) free cholesterol, and (3) one-tenth phospholipid concentration in plasma (in mg. per 100 cc.). Corresponding 90-per-cent normal range estimates are indicated.

ment appears to produce marked and sustained depression in all three curves of at least five of the six animals observed.

That this is statistically significant seems obvious. A single observation below the 90% normal range should be expected once in twenty random trials if normal causes of variation may be assumed unaffected by the treatment in question. The expectation of obtaining 5 or more such values in six independent trials is obviously much less, and may be accurately estimated by means of relations developed in the following section.

2. Significance Tests. Now consider as in *section* 1 another sample S' of n' values; $\{x'_{k'}\}$, $k' = 1, \cdots, n'$ (where $x'_i < x'_j$ if $i < j$), drawn at random from an infinite universe U' as was S from U; but where U' and U are not necessarily

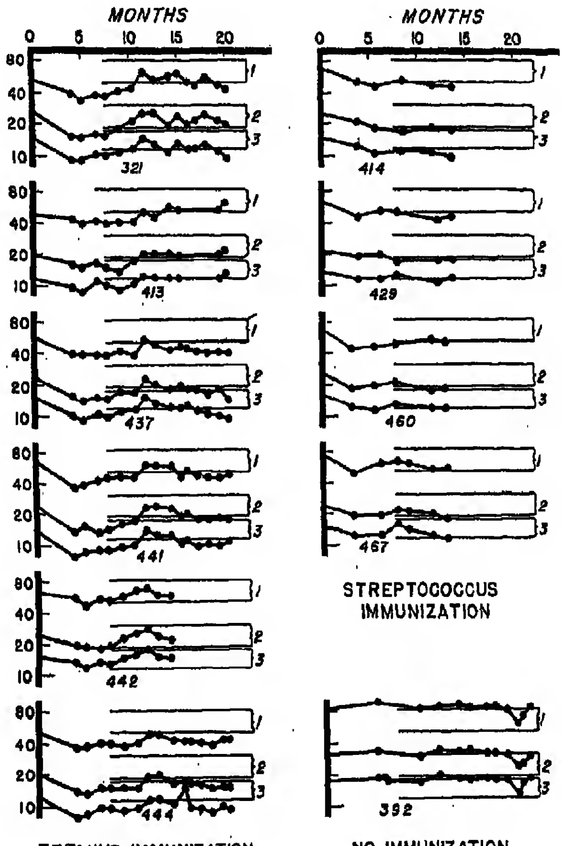

CHART 2. On each graph for horses receiving the indicated antigenic treatment and one untreated horse, the curves in descending order respectively represent (1) esterified cholesterol, (2) free cholesterol, and (3) one-tenth phospholipid concentration in plasma (in mg. per 100 cc.). Corresponding 90-per-cent normal range estimates are indicated.

the same universe. In like manner it may be shown that, if x' is drawn at random from U' and $p'_{k'}$ denotes $P(x' < x'_{k'})$, then

$$(8) \qquad P_{n'}(\phi' < p'_{k'} < \phi'') = \frac{(v + w + 1)!}{v!w!} \cdot \int_{\phi'}^{\phi''} p^v \cdot q^w \cdot dp$$

where $q = 1 - p$, $v = k' - 1$, $w = n' - k'$, and $0 \leqq \phi' \leqq \phi'' \leqq 1$.

The probabilities in (4) and (8) are independent, obviously, whether U' is the same as U or not. Accordingly, these relations make possible an evaluation

of $P(p_k < p'_{k'})$ under the circumstances where repeated sampling is applied to both the case of S and to that of S'. With this understanding, then

$$(9) \quad P(p_k < p'_{k'}) = \frac{(r+s+1)!(v+w+1)!}{r!\cdot s!\cdot v!\cdot w!} \cdot \int_0^1 p_0^r \cdot q_0^s \cdot dp_0 \cdot \int_{p_0}^1 p^v \cdot q^w \cdot dp,$$

where, as before, $r = k - 1$, $s = n - k$, $v = k' - 1$, $w = n' - k'$, $q \equiv 1 - p$, and $q_0 \equiv 1 - p_0$.

In a previous paper [4] a Ψ-function was defined as

$$(10) \qquad \Psi(r, s, r', s') \equiv \frac{\sum_{\alpha=0}^{r'} \binom{r + r' - \alpha}{r} \binom{s + s' + 1 + \alpha}{s}}{\binom{r + s + r' + s' + 2}{r + s + 1}}$$

for any four rational integers $r, s, r', s' \geqq 0$; and it was shown in detail that the right member of (9) is equal to $\Psi(r, s, v, w)$; whence we may write

$$(11) \qquad P(p_k < p'_{k'}) = \Psi(k - 1, n - k, k' - 1, n' - k').$$

Obviously, if U and U' are the same universe, then $p_k < p'_{k'}$ if and only if $x_k < x'_{k'}$, and then we have

$$(12) \qquad P(x_k < x'_{k'}) = \Psi(k - 1, n - k, k' - 1, n' - k')$$

in repeated random sampling applied to both sample types, S and S', respectively of n and of n' observations. In the paper just mentioned, and in another [5] the Ψ-function was further developed by extension of definition to include $\Psi(r, s, -1, s') \equiv 0$, and it was shown that

$$(13) \quad \Psi(r, s, r', s') \equiv \Psi(r, r', s, s') \equiv \Psi(s', r', s, r) \equiv 1 - \Psi(s, r, s', r').$$

Further demonstrations [5] included the relation,

$$(14) \qquad \Psi(r, s, r', s') \equiv \frac{\sum_{\alpha=0}^{\alpha \leqq s, r'} \binom{r + r' + 1}{r + 1 + \alpha} \binom{s + s' + 1}{s - \alpha}}{\binom{r + s + r' + s' + 2}{r + s + 1}},$$

which offers another form for calculation. The identities in (13) are particularly useful to facilitate calculation where one of the four arguments is small. A system for forming a table has also been developed [4, 5] in an economical form, but tabulation has been given only for the arguments not exceeding 5.

Now, in applying a test based on relation (12) or on that for the complementary probability, $P(x'_{k'} < x_k)$ which obviously, by (13), equals $\Psi(n - k, k - 1, n' - k', k' - 1)$, we may wish to exclude from the *normal* set of observations those values obtained from animals later given the treatment in question in the statistical significance test. The purpose would be to avoid violation of the condition of independent sampling required. In the case of the tetanus antigen treatment, we have an experience wherein 5 or more of 6 horses treated yield

values for a given plasma constituent less than the third in ascending order of magnitude (namely x_3) in our independent set of *normal* values. Here $n' = 6$, and $n = 62 - 6 = 56$. In accordance with the hypothesis that the treatment in question does not affect normal causes of variation in the plasma constituent under investigation we have $P(x'_{k'} < x_3)$ is $\Psi(53, 2, 6 - k', k' - 1)$. This is approximately $1.891(10)^{-5}$ for $k' = 5$, and $4.555(10)^{-7}$ for $k' = 6$. Obviously, a rule for establishing the value of k to be used in such tests should be fixed in advance without prejudice, as in the present case where we have taken $k \geqq g(n + 1) > k - 1$ for $g = 5\%$.

In the case of streptococcus immunization treatment, the corresponding test would have $n = 58$, $n' = 4$, $k = 3$, and $k' = 4, 3$, or 2; which would yield approximately $2.689(10)^{-5}$, $1.031(10)^{-3}$, or $1.817(10)^{-2}$, respectively for $P(x'_{k'} < x_3)$. Thus it appears that where such values are found (intuitively it would appear a fortiori if we compare instead with x_3 of the entire normal set of 62 values), their low magnitude appears to discredit the hypothesis that such discrepancies are ascribable to mere chance normal variation in the quantitative attribute investigated.

The tests proposed are free from any assumption concerning the form of the original distribution $f(x)$. The illustrative material is only a part of that presented with similar statistical treatment in the paper of Wadsworth and Hyman [3], which makes it apparent that the tests suggested here may be useful and powerful in analysis of biological and other experimental data. From a similar point of view, Hotelling and Pabst [6] developed tests of bi-variate correlation, and Milton Friedman has elaborated a multi-variate rank analysis [7], the tests being likewise free from any assumption about the form of the original distributions. In a previous paper [1] confidence ranges for the median are based similarly, employing relation (5) for the special case $p = \frac{1}{2}$.

Division of Laboratories and Research
New York State Department of Health
Albany, N. Y.

REFERENCES

[1] W. R. Thompson, *Annals of Mathematical Statistics*, Vol. 7 (1936), p. 122.
[2] *Tables of the Incomplete Beta-function*, edited by Karl Pearson, (Office of *Biometrika*, University College, London), 1934, p. 494.
[3] Augustus Wadsworth and L. W. Hyman, *Jour. Immunol.*, Vol. 35 (1938), p. 55.
[4] W. R. Thompson, *Biometrika*, Vol. 25 (1933), p. 285.
[5] W. R. Thompson, *American Journal of Mathematics*, Vol. 57 (1935), p. 450.
[6] H. Hotelling and M. R. Pabst, *Annals of Mathematical Statistics*, Vol. 7 (1936), p. 29.
[7] Milton Friedman, *Jour. Amer. Stat. Assoc.*, Vol. 32 (1937), p. 675.

THE COMPUTATION OF MOMENTS WITH THE USE OF CUMULATIVE TOTALS

By Paul S. Dwyer

1. **Introduction.** Various authors have shown how the moments of a frequency distribution may be computed from cumulated frequencies.[1] In order to make clear to the reader the type of technique under discussion there is presented an illustration which is, essentially, that used by Hardy, [2, p. 59]. The value $\Sigma f_x = 729$ is the last entry in column 4.

We use C_1^1 to denote the entry in column 4 which is opposite the smallest variate (or class mark if the distribution is grouped). Similarly C_2^1 is the entry above C_1^1, and C_1^2 the entry to the right of C_1^1, etc. In this notation the diagonal entries, the ones underscored in Table I, are C_1^1, C_2^2, C_3^3, C_4^4, C_5^5.

The moments[2] about the smallest variate can be expressed in terms of the cumulations of Table I in different ways. One method utilizes the diagonal entries and the differences of zero. Thus

$$\sum_0^6 x f_x = C_2^2 = 2916; \qquad \sum_0^6 x^2 f_x = C_2^2 + 2C_3^3 = 12636;$$

$$\sum_0^6 x^3 f_x = C_2^2 + 6C_3^3 + 6C_4^4 = 57996;$$

$$\sum_0^6 x^4 f_x = C_2^2 + 14C_3^3 + 36C_4^4 + 24C_5^5 = 278316, \text{ etc.}$$

A second method utilizes the entries in the next to the last row and the differences of zero. Thus

$$\sum_0^6 x f_x = C_2^2 = 2916; \qquad \sum_0^6 x^2 f_x = -C_2^2 + 2C_2^3 = 12636;$$

$$\sum_0^6 x^3 f_x = C_2^2 - 6C_2^3 + 6C_2 = 57996;$$

$$\sum_0^6 x^4 f_x = -C_2^2 + 14C_2^3 - 36C_2^4 + 24C_2^5 = 278316, \text{ etc.}$$

[1] The reader is referred to reference [1] . . . [15], at end of paper.

[2] It is to be noted that we are not talking about moments *per unit frequency*. We are using the term in the sense used for example by Whittaker and Robinson. See [20, p. 18].

A third method, which seems to have escaped previous attention, involves columnar entries and multipliers whose determination and properties are a chief concern of this paper. Thus

$$\sum_0^6 xf_x = C_2^2 = 2916; \qquad \sum_0^6 x^2f_x = C_2^3 + C_3^3 = 12636;$$

$$\sum_0^6 x^3f_x = C_2^4 + 4C_3^4 + C_4^4 = 57996;$$

$$\sum_0^6 x^4f_x = C_2^5 + 11C_3^5 + 11C_4^5 + C_5^5 = 278316, \text{ etc.}$$

It is possible also to obtain formulas when the cumulations are made from the smallest variate to the largest variate and, indeed, the whole theory of the present paper could be duplicated with such a theory of cumulation.

TABLE I

Successive Frequency Cumulations

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X	x	F_x	C^1	C^2	C^3	C^4	C^5
$a+6$	6	64	64	64	64	64	64
$a+5$	5	192	256	320	384	448	512
$a+4$	4	240	496	816	1200	1648	2160
$a+3$	3	160	656	1472	2672	4320	6480
$a+2$	2	60	716	2188	4860	9180	15660
$a+1$	1	12	728	2916	7776	16956	32616
a	0	1	729	3645	11421	28377	60993

It is possible to obtain the columnar formulas from the well known diagonal formulas. From the construction of Table I it is clear that

$$C_i^j = C_{i+1}^j + C_i^{j-1} \tag{1}$$

so that

$$\begin{gathered} C_2^2 = C_2^2; C_2^2 + 2C_3^2 = C_2^3 + C_3^3; \qquad C_2^2 + 6C_3^3 + 6C_4^4 = C_2^4 + 4C_3^4 + C_4^4; \\ C_2^2 + 14C_3^3 + 36C_4^4 + 24C_5^5 = C_2^5 + 11C_3^5 + 11C_4^5 + C_5^5. \end{gathered} \tag{2}$$

Formula (1) can be used similarly in deriving columnar formulas from row formulas, diagonal formulas from row formulas, etc.

The columnar method is here recommended as a useful substitute for the usual elementary method of computing moments. The many multiplications involved in the usual process are replaced by continued addition. The chief

disadvantage of the method is the continual recording, although this obstacle is surmounted with an adding machine equipped with a recording tape. The resulting moments are easily checked with an adaptation of Charlier's check, as is shown in section 8; and methods are given by which the multipliers are easily obtained. The method is also well adapted to the use of Hollerith machines.

The introduction of such columnar multipliers tends to give a different emphasis to the cumulative totals technique. The use of diagonal entries led logically to an emphasis upon factorial moments, while the columnar method tends to emphasize the more familiar power moments. The primary application here indicated is not to elaborate and specialized techniques, but rather to the simple, though often tedious, problem of the computation of power moments.

The aims of this paper are then:

(1) To show how moments may be computed from the columnar values of the successive cumulations,
(2) To discover the properties of the columnar multipliers,
(3) To present a general theory for computation of moments using cumulative totals.

2. **The Basic Cumulative Theorem.** The use of (1) is not satisfactory in getting precise formulas for the columnar multipliers so we derive the columnar cumulative theory directly from first principles. We first prove

THEOREM I. *Let x be any real number and let u_x be a real function of x which is 0 when $x < a$ and when $x > a + k$ and which is not infinite for $x = a, a + 1, a + 2, \cdots, a + k$. Let v_x be a real function of x and $\underline{v}_x$, called range v_x, a function such that $\underline{v}_x = v_x$ when $x = a, a + 1, \cdots, a + k$ and $\underline{v}_x = 0$ at all points outside the range a to $a + k$. If $\sum_{x}^{a+k} u_i$ is indicated by Cu_x and $\underline{v}_x - v_{x-1}$ by ∇v_x, $\underline{v}_x - \underline{v}_{x-1}$ by $\nabla \underline{v}_x$ then*

$$\sum_{a}^{a+k} u_x v_x = \sum_{a}^{a+k} u_x \underline{v}_x = \sum_{a}^{a+k} Cu_x \nabla \underline{v}_x . \tag{3}$$

The values u_x, v_x, Cu_x, $\nabla \underline{v}_x$ are presented in Table II.

The theorem is proved by forming

$$\sum_{a}^{a+k} Cu_x \nabla \underline{v}_x \equiv u_{a+k} v_{a+k} + \cdots + u_{a+i} v_{a+i} + \cdots + u_a v_a$$

$$\equiv \sum_{a}^{a+k} u_x v_x \equiv \sum_{a}^{a+k} u_x \underline{v}_x .$$

Theorem I can also be written as

$$\sum_{0}^{k} u_{a+x} v_{a+x} = \sum_{0}^{k} Cu_{a+x} \nabla \underline{v}_{a+x} \tag{4}$$

3. The Successive Cumulation Theorem.

THEOREM II. *If* $C^2u_x = C[Cu_x]$ *and* $\nabla^2\underline{v}_x = \nabla(\nabla\underline{v}_x)$, *etc., then*

$$\sum_a^{a+k} u_x v_x = \sum_a^{a+k} u_x \underline{v}_x = \sum_a^{a+k} C^{s+1}u_x\nabla^{s+1}\underline{v}_x.$$

This theorem follows readily from Theorem I. If

$$U_x = Cu_x \quad \text{and} \quad V_x = \nabla v_x, \qquad \text{then}$$

$$\sum_a^{a+k} u_x v_x = \sum_a^{a+k} u_x \underline{v}_x = \sum_a^{a+k} U_x \underline{V}_x = \sum_a^{a+k} CU_x\nabla\underline{V}_x = \sum_a^{a+k} C^2u_x\nabla^2\underline{v}_x.$$

This process can be extended as many times as desired so that

$$\sum_a^{a+k} u_x v_x = \sum_a^{a+k} u_x \underline{v}_x = \sum_a^{a+k} C^{s+1}u_x\nabla^{s+1}\underline{v}_x. \tag{5}$$

TABLE II

Values of x, u_x, v_x, Cu_x, *and* $\nabla\underline{v}_x$.

x	u_x	v_x	Cu_x	$\nabla\underline{v}_x$
$a+k$	u_{a+k}	v_{a+k}	u_{a+k}	$v_{a+k} - v_{a+k-1}$
$a+k-1$	u_{a+k-1}	v_{a+k-1}	$u_{a+k} + u_{a+k-1}$	$v_{a+k-1} - v_{a+k-2}$
....	....	....	....	....
$a+i$	u_{a+i}	v_{a+i}	$u_{a+k} + \cdots + u_{a+i}$	$v_{a+i} - v_{a+i-1}$
....	....	....	....	....
$a+1$	u_{a+1}	v_{a+1}	$u_{a+k} + \cdots + u_{a+i} + \cdots + u_{a+1}$	$v_{a+1} - v_a$
a	u_a	v_a	$u_{a+k} + \cdots + u_{a+i} + \cdots + u_{a+1} + u_a$	v_a

This can also be written as

$$\sum_0^k u_{a+x}v_{a+x} = \sum_0^k u_{a+x}\underline{v}_{a+x} = \sum_0^k C^{s+1}u_{a+x}\nabla^{s+1}\underline{v}_{a+x}. \tag{6}$$

In order to determine the values $\nabla^{s+1}\underline{v}_{a+x}$, $0 \leq x \leq k$, we note that

$$\nabla^{s+1}v_{a+x} = \sum_0^s (-1)^t \binom{s+1}{t} v_{a+x-t}, \tag{7}$$

so that

$$\nabla^{s+1}\underline{v}_{a+x} = \sum_0^s (-1)^t \binom{s+1}{t} \underline{v}_{a+x-t}. \tag{8}$$

We also know that, $x \leq k$

$$(9) \qquad \left.\begin{aligned} \underline{v}_{a+x-t} &= v_{a+x-t} && \text{when } t \leq x \\ \underline{v}_{a+x-t} &= 0 && \text{when } t > x \end{aligned}\right\}$$

so that

$$(10) \qquad \nabla^{s+1}\underline{v}_{a+x} = \sum_{0}^{x} (-1)^t \binom{s+1}{t} v_{a+x-t}, \qquad 0 \leq x \leq s$$

$$(11) \qquad \nabla^{s+1}\underline{v}_{a+x} = \sum_{0}^{s} (-1)^t \binom{s+1}{t} v_{a+x-t} = \nabla^{s+1} v_{a+x}, \qquad s < x \leq k.$$

The formula (6) can then be written

$$(12) \quad \sum_{a}^{a+k} u_x v_x = \sum_{0}^{k} u_{a+x} v_{a+x} = \sum_{0}^{s} C^{s+1} u_{a+x} \nabla^{s+1} \underline{v}_{a+x} + \sum_{s+1}^{k} C^{s+1} u_{a+x} \nabla^{s+1} v_{a+x}.$$

4. **Moments from the Cumulated Frequencies.** If $u_{a+x} = f_{a+x}$ and $v_{a+x} = (a + x)^s$, then (6) gives

$$(13) \qquad \sum_{0}^{k} (a + x)^s f_{a+x} = \sum_{0}^{k} C^{s+1} f_{a+x} \nabla^{s+1} \underline{(a + x)}^s.$$

A more useful formula, obtained from (12), is

$$(14) \qquad \sum_{0}^{k} (a + x)^s f_{a+x} = \sum_{0}^{s} C^{s+1} f_{a+x} \nabla^{s+1} \underline{(a + x)}^s,$$

since $\nabla^{s+1} (a + x)^s = 0$. We have then

THEOREM III. *The values of the s-th moments can be obtained from the last $s + 1$ entries of the $(s + 1)$st cumulation of the frequencies. The multipliers are the values*

$$(15) \qquad \nabla^{s+1}\underline{(a + x)}^s = \sum_{0}^{x} (-1)^t \binom{s+1}{t} (a + x - t)^s.$$

Cor. 1. When $a = 0$, i.e., when the moments are measured about the smallest variate, the multipliers are

$$(16) \qquad \nabla^{s+1}\underline{x}^s = \sum_{0}^{x} (-1)^t \binom{s+1}{t} (x - t)^s.$$

Cor. 2. When $a = 1$, the multipliers are

$$(17) \qquad \nabla^{s+1}\underline{(1 + x)}^s = \sum_{0}^{x} (-1)^t \binom{s+1}{t} (1 + x - t)^s.$$

Cor. 3. If the moments are measured about a fixed value, p, then the new smallest variate is $a - p = a'$ and the multipliers are $\nabla^{s+1} \underline{(a' + x)}^s$.

Cor. 4. If p is the mean, m, then $a' = a - m$. If in addition $a = 0$, then $a' = -m$ and the multipliers giving moments about the mean are $\nabla^{s+1}\,\underline{(x - m)}^s$. Now

$$m = \frac{\sum_0^k x f_x}{\sum_0^k f_x} = \frac{C_1^2 \nabla^2 \underline{0} + C_2^2 \nabla^2 \underline{1}}{C_1^1} = \frac{C_2^2}{C_1^1}.$$

It follows that the multipliers giving the moments about the mean are

$$\nabla^{s+1} \underline{\left(x - \frac{C_2^2}{C_1^1}\right)}^s. \tag{18}$$

It is to be noted that the moments about different points are obtained by applying different multipliers to the same cumulated frequencies.

5. Values of the multipliers. The values of the multipliers may be computed from (15). Thus $\nabla^3 \underline{(a + 1)}^2 = (a + 1)^2 - 3a^2 = -2a^2 + 2a + 1$. This becomes $2ab + 1$ when $1 - a$ is set equal to b. Values of the multipliers for the most common values of s and x are presented in Table III.

TABLE III

Values of $\nabla^{s+1}\,\underline{(a + x)}^s$

x \ s	0	1	2	3	4
4					b^4
3				b^3	$4b^3a + 6b^2 + 4b + 1$
2			b^2	$3b^2a + 3b + 1$	$6a^2b^2 + 12ab + 11$
1		b	$2ab + 1$	$3a^2b + 3a + 1$	$4a^3b + 6a^2 + 4a + 1$
0	1	a	a^2	a^3	a^4

When $a = 0, b = 1$ and the multipliers are 1; 0, 1, 1; 0, 1, 4, 1; 0, 1, 11, 11, 1; etc. as indicated in section 1. When $a = 1, b = 0$ and the multipliers are 1; 1, 0; 1, 4, 1, 0; 1, 11, 11, 1, 0; etc. When the moments are measured about a fixed point, p, it is only necessary to compute $a' = a - p$ and to use a' for a and $b' = 1 - a'$ for b in Table III.

We illustrate the use of the multipliers by application to the problem of Table I. The moments about the smallest variate are computed in section 1. The moments, when $a = 1$ are $\sum_0^6 (x + 1)f_x = C_1^2 = 3654$; $\sum_0^6 (x + 1)^2 f_x =$

$C_1^3 + C_2^3 = 19197; \sum_0^6 (x+1)^3 f_x = C_2^4 + 4C_3^4 + C_4^4 = 105381; \sum_0^6 (x+1)^4 f_x =$ $C_2^5 + 11C_3^5 + 11C_4^5 + C_5^5 = 598509.$

The moments about the mean are found by forming $\frac{C_2^2}{C_1^1} = \frac{2916}{729} = 4$. Then $a = -4$ and the multipliers are 1; -4, 5; 16, -39, 25; -64, 229, -284, 125; 256, -1199, 2171, -1829, 625; etc. so that $\sum_0^6 \bar{x} f_x = 0; \sum_0^6 \bar{x}^2 f_x = 972; \sum_0^6 \bar{x}^3 f_x = -324; \sum_0^6 \bar{x}^4 f_x = 3564.$

Since the values of $\nabla^{s+1}(x - C_2^2/C_1^1)^s$ are expressible in terms of C_1^1 and C_2^2, it follows that the values of $\sum_0^k \bar{x}^s f_x$ are expressible in terms of cumulations. For example a formula for the second moment about the mean, which is essentially one given by Whittaker and Robinson [7, p. 193] is

$$\sum_a^{a+k} \bar{x}^2 f_x = C_2^2 + 2C_3^3 - \frac{(C_2^2)^2}{C_1^1}. \tag{19}$$

However the general method described above, supplemented with the techniques of succeeding sections, is preferred to the development and use of such formulas.

6. Recursion Property of the Multipliers. It is not readily apparent from Table III how the multipliers of the $(s+1)$-th cumulations can be obtained from the multipliers of the s-th cumulations. It is possible to establish a recursion formula which is useful for this purpose. Now, $a \leq x \leq s$,

$$\nabla^{s+1}\underline{(a+x)}^s = (a+x)^s + \sum_1^x (-1)^t \binom{s+1}{t} (a+x-t)^s$$

$$(a+x)\nabla^s\underline{(a+x)}^{s-1} = (a+x)^s + \sum_1^x (-1)^t \binom{s}{t} (a+x-t)^{s-1}(a+x)$$

$$(s+1-a-x)\nabla^s\underline{(a+x-1)}^{s-1}$$

$$= \sum_1^x (-1)^{t-1} \binom{s}{t-1} (a+x-t)^{s-1}(s+1-a-x)$$

and since

$$\binom{s}{t}(a+x) - \binom{s}{t-1}(s+1-a-x) = \binom{s+1}{t}(a+x-t)$$

it follows that

$$\Delta^{s+1}\underline{(a+x)}^s = (a+x)\nabla^s\underline{(a+x)}^{s-1} + (s+1-a-x)\nabla^s\underline{(a+x-1)}^{s-1}. \tag{20}$$

When $a = 0$ we have

$$(21) \qquad \nabla^{s+1}\underline{x}^{s} = x\nabla^{s}\underline{x}^{s-1} + (s + 1 - x)\nabla^{s}\underline{(x - 1)}^{s-1}.$$

Formulas (20) and (21), though somewhat formidable in appearance, are easy to apply. Thus $\nabla^{3}\underline{(a + 2)}^{2} = (a + 2)\nabla^{2}\underline{(a + 2)} + (1 - a)\nabla^{2}\underline{(a + 1)}$. The recursion formula is especially useful in building up tables of multipliers. The following form is recommended:

As successive columnar headings use the values a, $a + 1$, $a + 2$, etc. and as successive row headings use $1 - a$, $2 - a$, $3 - a$, etc. Then $\nabla\underline{a}^{0} = 1$ is placed in the upper left cell, $\nabla^{2}a$ directly below $\nabla\underline{a}^{0}$, $\nabla^{2}\underline{a + 1}$ to the right of $\nabla^{2}\underline{a}^{0}$, etc. The values of $\nabla^{3}\underline{(a + x)}^{2}$ are placed in the next diagonal, etc. If this process is continued the entry $\nabla^{s+1}\underline{(a + x)}^{s}$ will have the entry $\nabla^{s}\underline{(a + x)}^{s-1}$ directly above it and the entry $\nabla^{s}\underline{(a + x - 1)}^{s-1}$ on its left. Also the columnar heading is $a + x$ and the row heading $s + 1 - a - x$ so that any entry is obtained by adding the product of the entry above it and the columnar heading to the product of the entry to the left and the row heading. The values of $\nabla^{s+1}\underline{x}^{s}$ are obtained by placing $a = 0$. They are presented, in Table IV, through $s = 8$.

TABLE IV

Values of $\nabla^{s+1}\underline{x}^{s}$

$s+1-x$ \ x	0	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1	1
2	0	1	4	11	26	57	120	247	
3	0	1	11	66	302	1191	4293		
4	0	1	26	302	2416	15619			
5	0	1	57	1191	15619				
6	0	1	120	4293					
7	0	1	247						
8	0	1							

The table is easily extended to higher values of s. If a table of values of $\nabla^{s+1}\underline{(x + 1)}^{s}$ is constructed, it will be found to be like Table IV with columns and rows interchanged. Hence the values of $\nabla^{s+1}\underline{(x + 1)}^{s}$ are obtained from Table

IV by reading the multipliers down the diagonal. Thus the values $\nabla^{s}\underline{(x+1)}^{2}$ are 1, 4, 1, 0, etc.

The ease with which the multipliers may be computed is illustrated with $a = -4$. In this case we have

TABLE V

Values of $\nabla^{s+1}\,\underline{(x+a)}^{s}$ *with* $a = -4$

$s+1-a-x$ \\ $a+x$	-4	-3	-2	-1	0
5	1	5	25	125	625
6	-4	-39	-284	-1829	
7	16	229	2171		
8	-64	-1199			
9	256				

These values agree with those computed more laboriously in section 5.

7. **Value of** $\sum_0^k \nabla^{s+1}\underline{(x+a)}^{s}$. It is to be noted in Tables III, IV, V that the sum of the entries in the diagonal having $s + 1$ terms is $s!$ This is generally true and results from the fact that

$$\sum_0^k \nabla^{s+1}\underline{(x+a)}^{s} = \sum_0^s \nabla^{s+1}\underline{(x+a)}^{s} = s! \tag{22}$$

In obtaining the values of $\sum_0^k \nabla^{s+2}\underline{(x+a)}^{s+1}$ from the value of $\sum_0^k \nabla^{s+1}\underline{(x+a)}^{s}$ it is noted that $\nabla^{s+1}\underline{(x+a)}^{s}$ is used but twice. Once it is multiplied by $a + x$ and once by $s + 1 - a - x$ so that the net result is a multiplication by $s + 1$. It follows that $\sum_0^k \nabla^{s+2}\underline{(x+a)}^{s+1} = (s+1)\sum_0^k \nabla^{s+1}\underline{(x+a)}^{s}$ and since $\sum_0^k \nabla^{2}\underline{(x+1)}$ $= 1$, $\sum_0^k \nabla^{3}\underline{(x+a)}^{2} = 2!$ so that in general $\sum_0^k \nabla^{s+1}\underline{(x+a)}^{s} = s!$

This property is useful in checking the values of the computed multipliers.

8. **The adaptation of the Charlier check.** An adaptation of the Charlier check serves as an excellent check for the computed moments. It is recalled that the Charlier check gives

$$\sum_a^{a+k} (x+1)^s f_x = \sum_{t=0}^{s} \binom{s}{t} \sum_{x=a}^{a+k} x^{s-t} f_x. \tag{23}$$

The components of the right hand member are computed by cumulative totals as indicated above. The left hand member is obtained by applying different multipliers to the same cumulated frequencies. Thus $\sum_{a}^{a+k} (x+1)^s f_x = \sum_{0}^{k} (x + a + 1)^s f_{x+a}$ and the multipliers of the cumulated frequencies are $\nabla^{s+1}\underline{(x + a')^s}$ where $a' = a + 1$. If $a = 0$ the Charlier check multipliers are the values $\nabla^{s+1}\underline{(x + 1)^s}$ which can be read from Table IV. For example $\sum_{0}^{6} (x+1)^4 f_x = C_1^5 + 11C_2^5 + 11C_3^5 + C_4^5 = 598509$ and this checks with $\sum_0^6 x^4 f_x + 4\sum_0^6 x^3 f_x + 6\sum_0^6 x^2 f_x + 4\sum_0^6 x f_x + \sum_0^6 f_x$.

9. **Application to factorial moments.** When $u_x = f_x$, $v_x = x^{(s)} = x(x-1)(x-2)\cdots(x-s+1)$

$$\sum_0^k x^{(s)} f_x = \sum_0^k C^{s+1} f_x \nabla^{s+1} \underline{x}^{(s)}$$

and since $\nabla^{s+1}\underline{x}^{(s)}$ is 0 when $s < x \leq k$, is $s!$ when $s = x$, is 0 when $0 \leq x < s$,

$$\sum_0^k x^{(s)} f_x = \sum_s^k x^{(s)} f_x = s!\, C_{s+1}^{s+1}. \tag{24}$$

It follows that the underscored terms of Table I, when multiplied by $s!$, give the factorial moments. Factorial moments, first used by Sheppard [4], have since come into prominence largely because of this ease of computation.

The coefficients of $(a + b)^x$ are $1, x, \frac{x(x-1)}{2!}, \cdots, \frac{x^{(s)}}{s!}, \cdots$. If we define the binomial moment by $B_s = \sum_0^k \frac{x^{(s)}}{s!} f_x$ [6, p. 278] then $B_s = \frac{1}{s!}\sum_0^k x^{(s)} f_x = C_{s+1}^{s+1}$.

It is also possible to show that the entries under the main diagonal are binomial moments. In Table I, for example, we let $a = 1$ and add the additional row $a = 0$ with 0 frequency. Then $C_1^1 = 729$, $C_2^1 = 729$, $C_1^2 = 729 + 3645 = 4374$, etc. The new diagonal terms are directly under the old diagonal terms and give $B_{s,1} = \sum_1^7 x^{(s)} f_x = \sum_0^6 (x+1)^{(s)} f_x$. In general the terms $B_{s,l}$ are given l rows below the terms B_s and the factorial moments are $s!\,B_{s,l}$. Then

$$F_{s,l} = s!\, C_{s+1-l}^{s+1}. \tag{25}$$

For example in the problem of Table I, $F_{4,3} = \sum_3^9 x^{(4)} f_x = 4!\,C_2^5 = 782{,}784$. The method is especially adapted to the use of Hollerith machines, for positive integral values of l, since it is only necessary to have the machine continue its cumulation.

10. The cumulations of xf_x. It is possible to use the cumulations of xf_x in securing the values of the moments. Now

$$\sum_{a}^{a+k} x^{s+1}f_x = \sum_{0}^{k} (x+a)^{s+1}f_{x+a} = \sum_{0}^{k} (x+a)f_{x+a}(x+a)^s$$

(26)

$$= \sum_{0}^{s} C^{s+1}(x+a)f_{x+a}\nabla^{s+1}\underline{(x+a)^s}.$$

When $a = 0$, (26) becomes

(27) $$\sum_{0}^{k} x^{s+1}f_x = \sum_{0}^{s} C^{s+1}xf_x\nabla^{s+1}\underline{x}^s.$$

We compute the cumulations of xf for the problem of Table I. These are given in Table VI.

TABLE VI

Cumulations of xf_x

x	f_x	xf_x	C^1	C^2	C^3	C^4
6	64	384	384	384	384	384
5	192	960	1344	1728	2112	2496
4	240	960	2304	4032	6144	8640
3	160	480	2784	6816	12960	21600
2	60	120	2904	9720	22680	44280
1	12	12	2916	12636	35316	79596
0	1	1	2916	15552	50868	130464

so that

$$\sum_{0}^{6} xf_x = 2916; \quad \sum_{0}^{6} x^2f_x = 12636; \quad \sum_{0}^{6} x^3f_x = 35316 + 22680 = 57996;$$

$$\sum_{0}^{6} x^4f_x = 79596 + 4(44280) + 21600 = 278316.$$

In getting moments about the mean from the cumulations of xf_x, the following method is recommended.

(28) $$\sum_{0}^{k} \bar{x}^{s+1}f_x = \sum_{0}^{k} \bar{x}^s(x-m)f_x = \sum_{0}^{k} \bar{x}^s xf_x - m\sum_{0}^{k} \bar{x}^s f_x.$$

and

(29) $$\sum_{0}^{k} \bar{x}^s xf_x = \sum_{0}^{k} C^{s+1}(xf_x)\nabla^{s+1}(x-m)^s.$$

When $s = 1$, (28) gives $\sum_{0}^{k} \bar{x}^2f_x = \sum_{0}^{k} \bar{x}xf_x - m\sum_{0}^{k} \bar{x}f_x$ and

$$\sum_0^k \bar{x}^2 f_x = \sum_0^k \bar{x} x f_x. \tag{30}$$

In the illustrative problem $a = -4$ so that

$$\sum_0^6 \bar{x} x f_x = -4(15552) + 5(12636) = 972$$

$$\sum_0^6 \bar{x}^2 x f_x = 16(50868) - 39(35316) + 25(22680) = 3564$$

$$\sum_0^6 \bar{x}^3 x f_x = 2268$$

and

$$\sum_0^6 \bar{x}^2 f_x = 972; \qquad \sum \bar{x}^3 f_x = 3564 - 4(972) = -324; \qquad \sum \bar{x}^4 f_x = 3564.$$

Formula (30) is of note since it permits the determination of $\sum_0^k \bar{x}^2 f_x$ directly from the cumulations of xf_x.

The factorial moments are also related to the cumulations of xf_x. Thus

$$\sum_0^k x^{(s)} f_x = \sum_0^k (x-1)^{(s-1)} x f_x = \sum_0^k C^s(xf_x) \nabla^s \underline{(x-1)^{(s-1)}} \tag{31}$$

which results in $\sum_0^k x^{(s)} f_x = (s-1)! C_s^s(xf_x)$.

It follows that

$$C_s^s(xf_x) = sC_{s+1}^{s+1}(f_x).$$

For example, the underscored terms of Table VI are respectively 1, 2, 3, 4 times underscored terms of Table I.

In general the cumulations of xf_x, rather than of f_x, are recommended since $C(xf_x)$ can be computed and recorded almost as quickly as $C(f_x)$, since one less cumulation is needed to obtain a specific moment, and since the multipliers needed to get a specific moment are smaller. A technique based on the cumulations of xf_x is especially adapted to the use of Hollerith machines. Let us take x_x to represent the sum of the x's for all items in the distribution having the same value of x. Then $xf_x = x_x$ and we have

$$\sum_a^{a+k} x^s f_x = \sum_a^{a+k} x^{s-1} x_x = \sum_a^{a+k} C^s(x_x) \nabla^s \underline{(x^{s-1})}. \tag{32}$$

If the cards are sorted for x and the tabulator is wired to print cumulative totals each time x changes, the recording tape gives the successive values of $C(x_x)$. (Care must be taken that there are no blank values of x.)

If a summary punch is available, these cumulations are punched on cards as

they are cumulated and these summary cards are used in getting higher cumulations.

If no summary punch is available, it is possible to obtain $\sum x^2 f_x$ by the application of Theorem I. Thus

$$\sum_a^{a+k} x^2 f_x = \sum_a^{a+k} x x_x = \sum_a^{a+k} C(x_x)\nabla(\underline{x}),$$

and since $\nabla(\underline{x}) = a$ when $x = a$ and $\nabla(\underline{x}) = 1$ when $x > a$, it follows that $\sum_a^{a+k} x^2 f_x$ can be obtained by adding the entries above the last and then adding the last entry multiplied by a. This is essentially the Mendenhall-Warren-Hollerith method of getting $\sum x^2 f_x$ [9, p. 27].

In case $a = 0$ the technique amounts simply to adding all the entries above the bottom one.

The value $\sum x^3 f_x$ can be obtained similarly from the first order cumulations. Thus

$$\sum_a^{a+k} x^3 f_x = \sum_a^{a+k} x^2 x_x = \sum_a^{a+k} C(x_x)\nabla(\underline{x^2}) \tag{33}$$

and since $\nabla(\underline{x^2}) = a^2$ when $x = a$, $\nabla(\underline{x^2}) = 2x - 1$ when $x > a$, it follows that

$$\sum_a^{a+k} x^3 f_x = a^2 C_1^1(x_x) + \sum_{a+1}^{a+k} C(x_x)(2x - 1). \tag{34}$$

When $a = 0$, (34) becomes

$$\sum_0^k x^3 f_x = \sum_1^k C(x_x)(2x - 1) \tag{35}$$

so that the multipliers are the successive odd integers. Thus from the first order cumulations of Table VI we have

$$\sum_0^6 x f_x = 2916; \quad \sum_0^6 x^2 f_x = 12636; \quad \sum_0^6 x^3 f_x = 57996.$$

The cumulative method can also be applied to the method of digiting [17, p. 425].

It is also possible to obtain the moments from the cumulations of $x^2 f_x$, $x^3 f_x$, etc., since

$$\sum_a^{a+k} x^{s+2} f_x = \sum_a^{a+k} x^s x^2 f_x = \sum_a^{a+k} C^{s+1}(x^2 f_x)\nabla^{s+1}(\underline{x^s})$$

$$\sum_a^{a+k} x^{s+3} f_x = \sum_a^{a+k} x^s x^3 f_x = \sum_a^{a+k} C^{s+1}(x^3 f_x)\nabla^{s+1}(\underline{x^s}).$$

but the cumulations of $x f_x$ are preferable for most purposes. The Charlier check works in all cases. It should be noted that the indicated Hollerith technique

demands only the customary tabulator and not the expensive, time consuming, card punching, multiplier, [16].

11. **Product Moments. Correlation.** It is possible to apply the cumulative technique in getting product moments involving two variables. If we let y_x be the sum of all the values of y having the same value of x, then

$$\sum x^s y f_{xy} = \sum_{a}^{a+k} y_x x^s = \sum_{a}^{a+s} C^{s+1}(y_x)\nabla^{s+1}(\underline{x}^s) \tag{36}$$

so that the multipliers are the same as those previously used. When Hollerith machines are used, it is only necessary to sort the cards for x and to wire the machine to give cumulations on variables x, y, z, etc. If the machine is adjusted to take totals with each change in x, the tape records simultaneously the values of $C(x_x)$, $C(y_x)$, $C(z_x)$, etc. With a summary punch it is possible to form successive cumulations easily. The values Σx^{s+1}, $\Sigma x^s y$, $\Sigma x^s z$, etc. are then obtained by applying the multipliers. When $s = 1$, (36) becomes

$$\sum xyf_{xy} = \sum_{a}^{a+k} C^2(y_x)\nabla^2(\underline{x}) \tag{37}$$

so that the multipliers are a, $1 - a$, 0, 0, etc. When $a = 0$, the multipliers are 0, 1, 0, 0, etc. and when $a = 1$, they are 1, 0, 0, etc.

When no summary punch is available, it is necessary to obtain the values of the moments from the first order cumulations. Using Theorem I

$$\sum xyf_{xy} = \sum_{a}^{a+k} C(y_x)\nabla(\underline{x}) = aC_1^1(y_x) + \sum_{a+1}^{a+s} C(y_x). \tag{38}$$

This formula serves as the basis of the Mendenhall-Warren-Hollerith Correlation Method, [9, p. 27].

It can be shown in similar fashion that

$$\sum x^2 yf_{xy} = a^2C_1^1 + \sum_{a+1}^{a+s} C(y_x)(2x - 1) \tag{39}$$

and when $a = 0$

$$\sum x^2 yf_{xy} = \sum_{1}^{s} C(y_x)(2x - 1). \tag{40}$$

The method is also adapted to the common problem of finding correlation coefficients from grouped data when Hollerith machines are not available and this method is recommended for the determination of these coefficients.

An illustration is presented in Table VII which shows the correlation existing between college first semester average, X, and preparatory school average, Y, for 1126 students entering the College of Literature, Science and the Arts of the University of Michigan in 1928. The coded values of X and Y are indicated by x and y and are positive integers beginning with 0. The coded values are given

in descending order beginning with the upper left hand corner of the chart. The values of the cumulations are placed at the right hand side and at the bottom of the chart.

TABLE VII

Correlation with cumulative totals

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
$Y \backslash X$			4.00	3.99 3.50–	3.49 3.00–	2.99 2.50–	2.49 2.00–	1.99 1.50–	1.49 1.00–	.99 .50–	.49 .00–		
	$y \backslash x$		8	7	6	5	4	3	2	1	0		
		$f_y \backslash f_x$	13	50	107	220	341	179	121	60	35	Cx_y	Cy_y
4.00	6	18	5	2	5	5	1					113	108
3.99 3.50–	5	106	2	19	29	27	20	7		1	1	673	638
3.49 3.00–	4	178	3	12	35	53	44	18	6	5	2	1503	1350
2.99 2.50–	3	270	3	10	20	55	103	33	27	11	8	2568	2160
2.49 2.00–	2	330		6	11	54	114	67	46	19	13	3714	2820
1.99 1.50–	1	173		1	5	19	45	44	34	18	7	4244	2993
1.49 1.00–	0	51			2	7	14	10	8	6	4	4399	2993
		Cy_x	61	259	661	1330	2194	2578	2809	2923	2993	12815	10069
		Cx_x	104	454	1096	2196	3560	4097	4339	4399	4399	20245	1126

The lower right hand corner has the entries

$$\left.\begin{matrix} & \sum x & \sum y \\ \sum y & \sum xy & \sum y^2 \\ \sum x & \sum x^2 & \sum f = N \end{matrix}\right\}$$ where $\sum xy$, $\sum y$, and $\sum x$ are obtained by adding the cumulations in the columns or rows involved.

The values $C(y_y)$ are easily computed from columns (2) and (3). The values of $C(x_y)$ are computed by forming the cumulated product of the row frequency and x. The values are recorded when the products contributed by a given row have been computed. The values $C(y_x)$ and $C(x_x)$ are obtained similarly.

The value of r is easily obtained from the lower right hand entries. The value $A_{x,y} = N\Sigma xy - (\Sigma x)(\Sigma y)$ is obtained from diagonal entries, $A_{x,x} = N\Sigma x^2 - (\Sigma x)^2$

is obtained from entries in the last row, $A_{y,y} = N\Sigma y^2 - (\Sigma y)^2$ is obtained from the last column, and $r = \dfrac{A_{x,y}}{\sqrt{A_{x,x}A_{y,y}}}$ is easily computed. In the above problem $r = .441$.

The values M_x, M_y, σ_x, σ_y are also easily obtained from the lower right hand entries. The successive steps are indicated by the form

	Σx	Σy			M_y
Σy	Σxy	Σy^2			
Σx	Σx^2	N	$A_{x,x}$		
		$A_{y,y}$	$A_{x,y}$	$\sqrt{A_{y,y}}$	σ_y
			$\sqrt{A_{x,x}}$	$\sqrt{A_{x,x}A_{y,y}}$	
M_x			σ_x		r

Recent methods of applying cumulative totals theory to correlation are found in references [9], [14], [17], [18], [19].

The third order moments are obtained by multiplying the entries of $C(x_y)$, $C(y_y)$, $C(x_x)$, $C(y_x)$ by 1, 3, 5, etc. as indicated by (40). Thus $\Sigma x^3 f_x = 4399 + 3(4339) +$ etc. $= 102,103$; $\Sigma x^2 y f_{xy} = 63121$; $\Sigma xy^2 f_{xy} = 46047$; $\Sigma y^3 f_y = 38,633$. It is hence possible to compute the skewness of each marginal distribution from Table VII. See also [18, p. 657].

12. Conclusion. This paper presents an outline of the computation of moments with the use of cumulative totals and columnar multipliers. Basic general theorems are derived and applications are made to one variable and two variable distributions both with and without punched card equipment. The formulas assume that the distance between successive variates (or class marks) is unity, but the reader should have no trouble in adapting the formulas to more general problems.

In the interest of brevity the development is limited to the descending cumulations. It is possible to parallel the development here by deriving formulas in terms of ascending cumulations. It is also possible to work out formulas showing relations between columnar, row, and diagonal multipliers. There are other applications such as to the evaluation of $\sum_1^n x^s$, which are of interest. It is possible also that applications may be found for the general theory of sections 2 and 3 which do not demand that v_x be a power function.

THE UNIVERSITY OF MICHIGAN.

REFERENCES

[1] G. F. Lipps, "Die Theorie der Kollectivgegenstände," *Philosophische Studien* (Wundt Editor), Vol. 17, (1901) pp. 467-575.

[2] G. F. Hardy, *Theory of the Construction of Tables of Mortality*, pp. 59-62 and 124-128.

[3] W. P. Elderton, *Frequency Curves and Correlation*, pp. 19-23.

[4] W. F. Sheppard, "Factorial Moments in Terms of Sums or Differences," *Proc. of London Math. Society*, 2, Vol. 13, (1913) pp. 81-96.
Also, "Fitting Polynomials by the Method of Least Squares," ibid. pp. 97-108.

[5] J. Steffensen, "Factorial Moments and Discontinuous Frequency Functions," *Skandinavisk Aktuarietidskrift*, 6, (1923) pp. 73-89.

[6] J. Steffensen, *Interpolation*, pp. 93-104.

[7] E. T. Whittaker and G. Robinson, *Calculus of Observations*, pp. 191-194.

[8] R. Frisch, "Sur le calcul numérique des moments ordinaires et des moments composés d'une distribution statistique," *Skandinavisk Aktuarietidskrift*, Vol. 10, (1927) pp. 81-91.

[9] R. M. Mendenhall and R. Warren, "The Mendenhall-Warren-Hollerith Correlation Method," *Columbia University Statistical Bureau Document* No. 1.1929, Columbia University, New York, 43 pp.

[10] R. M. Mendenhall and R. Warren, "Computing Statistical Coefficients from Punched Cards," *Jour. of Ed. Psy.*, Vol. 21, (1930) pp. 53-62.

[11] C. Jordan, "Approximation and Graduation According to the Principle of Least Squares by Orthogonal Polynomials," *Annals of Math. Stat.*, 3, (1932) pp. 257-358.

[12] A. C. Aitkin, "On the Graduation of Data by the Orthogonal Polynomials of Least Squares," *Proc. of Roy. Soc. of Edin.*, Vol. 53, pp. 54-78.

[13] A. C. Aitkin, "On Fitting Polynomials to Weighted Data by Least Squares," *Proc. of Roy. Soc. of Edin.*, Vol. 54, (1933-34) pp. 1-11.

[14] Chen-nan Li, "Summation Method of Fitting Parabolic Curves and Calculating Linear and Curvilinear Coefficients on a Scatter Diagram," *Jour. of Am. Stat. Assn.*, 29, (1934) pp. 405-409.

[15] M. Sasuly, *Trend Analysis of Statistics*, Chap. VIII. Also page 5.

[16] H. C. Carver, "Uses of the Automatic Multiplying Punch"; *Punched Card Method in Colleges and Universities*, pp. 417-422.

[17] A. E. Brandt, "Uses of the Progressive Digit Method"; *Punched Card Method in Colleges and Universities*, pp. 423-436.

[18] P. S. Dwyer and A. D. Meacham, "The Preparation of Correlation Tables on a Tabulator Equipped with Digit Selection," *Jour. Am. Stat. Assn.*, Vol. 32, (1937) pp. 654-662.

[19] W. N. Durost and H. M. Walker, *Durost-Walker Correlation Chart* World Book Co. N. Y. (1938).

[20] H. L. Rietz, *Mathematical Statistics*, 1927.

A NOTE ON THE DERIVATION OF FORMULAE FOR MULTIPLE AND PARTIAL CORRELATION*

By Louis Guttman

1. **Multiple Correlation.** Let the measurements of N individuals on each of the n variables $x_1, x_2, \cdots, x_k, \cdots, x_n$, be expressed as relative deviates; that is, such that

$$\Sigma x_k = 0, \qquad \Sigma x_k^2 = N, \qquad k = 1, 2, 3, \cdots, n,$$

where the summations extend over the N individuals.

If values of λ_k are determined so that

$$\Sigma(x_1 - \lambda_2 x_2 - \lambda_3 x_3 - \cdots - \lambda_n x_n)^2 \text{ is a minimum,}$$

and if we let

$$X_1 = \lambda_2 x_2 + \lambda_3 x_3 + \cdots + \lambda_n x_n, \tag{1}$$

then the multiple correlation coefficient, obtained from the regression of x_1 on the remaining $n - 1$ variables, is defined as

$$r_{1.234\cdots n} = r_{x_1 X_1}.$$

The square of the standard error of estimate of x_1 on the remaining $n - 1$ variables is defined as

$$\sigma^2_{1.234\cdots n} = \frac{1}{N}\Sigma(x_1 - X_1)^2.$$

The minimizing values for λ_k are obtained from the normal equations

$$\Sigma(x_1 - X_1)x_k = 0, \qquad k = 2, 3, \cdots, n. \tag{2}$$

which may be written in expanded notation as,

$$\begin{array}{l}
\lambda_2 + r_{23}\lambda_3 + r_{24}\lambda_4 + \cdots + r_{2n}\lambda_n = r_{12} \\
r_{32}\lambda_2 + \lambda_3 + r_{34}\lambda_4 + \cdots + r_{3n}\lambda_n = r_{13} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
r_{n2}\lambda_2 + r_{n3}\lambda_3 + r_{n4}\lambda_4 + \cdots + \lambda_n = r_{1n}
\end{array}$$

where $r_{jk} = \dfrac{1}{N}\Sigma x_j x_k = r_{kj}, r_{jj} = 1.$

---

* The notions involved in this demonstration are certainly well-known. However, the directness and simplicity of the derivations may lend some merit to their exhibition. The writer is indebted to Professor Dunham Jackson for useful advice.

From Cramer's rule it is seen that

$$\lambda_k = -\frac{R_{1k}}{R_{11}}, \quad \text{if} \quad k \neq 1, \; R_{11} \neq 0,$$

where R_{jk} is the cofactor of r_{jk} (or of r_{kj}) in the symmetric determinant

$$R = \begin{vmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & r_{jk} & \cdots \\ r_{n1} & r_{n2} & r_{n3} & \cdots & r_{nn} \end{vmatrix}.$$

Summing both sides of (1) over the N individuals shows that $\Sigma X_1 = 0$, so that the variance of X_1 is

$$\sigma_{X_1}^2 = \frac{1}{N} \Sigma X_1^2.$$

From (2), the residual $(x_1 - X_1)$ is orthogonal to each of the x_k except x_1; therefore the residual is orthogonal to any linear combination of these x_k and in particular to X_1; that is,

(3) $$\Sigma(x_1 - X_1)X_1 = 0,$$

or

$$\sigma_{X_1} r_{x_1 X_1} = \sigma_{X_1}^2$$

and therefore

(4) $$r_{x_1 X_1} = \sigma_{X_1}.$$

Multiplying both sides of (1) by $\frac{x_1}{N}$ and summing over the individuals, we get:

$$\begin{aligned} \sigma_{X_1} r_{x_1 X_1} &= r_{12}\lambda_2 + r_{13}\lambda_3 + \cdots + r_{1n}\lambda_n \\ &= -\frac{1}{R_{11}} (r_{12}R_{12} + r_{13}R_{13} + \cdots + r_{1n}R_{1n}) \\ &= 1 - \frac{R}{R_{11}}. \end{aligned}$$

From (4) then,

$$r_{1.234\cdots n}^2 = 1 - \frac{R}{R_{11}}.$$

It is clear that in general

$$r_{k.123\cdots,k-1,k+1,\cdots n}^2 = 1 - \frac{R}{R_{kk}}.$$

To find the standard error of estimate, expand

$$\frac{1}{N}\Sigma(x_1 - X_1)^2 = 1 - 2\sigma_{X_1} r_{x_1 X_1} + \sigma^2_{X_1}$$

$$= 1 - r^2_{x_1 X_1}$$

$$= \frac{R}{R_{11}}.$$

In general, when $\sigma_k = 1$,

$$\sigma^2_{k.123\cdots,k-1,k+1,\cdots n} = \frac{R}{R_{kk}}. \tag{5}$$

2. **Partial Correlation.** If values of μ_k and ν_k are determined so that

$$\Sigma(x_1 - \mu_3 x_3 - \mu_4 x_4 - \cdots - \mu_n x_n)^2 \text{ is a minimum}$$

and $$\Sigma(x_2 - \nu_3 x_3 - \nu_4 x_4 - \cdots - \nu_n x_n)^2 \text{ is a minimum,}$$

and if we let

$$\begin{aligned} Y_1 &= \mu_3 x_3 + \mu_4 x_4 + \cdots + \mu_n x_n \\ Y_2 &= \nu_3 x_3 + \nu_4 x_4 + \cdots + \nu_n x_n, \end{aligned} \tag{6}$$

then the partial correlation coefficient between x_1 and x_2, holding the remaining $n - 2$ variables constant, is defined as

$$r_{12.34\cdots n} = r_{(x_1 - Y_1)(x_2 - Y_2)};$$

and since $\Sigma(x_k - Y_k) = 0$,

$$r_{12.34\cdots n} = \frac{\frac{1}{N}\Sigma(x_1 - Y_1)(x_2 - Y_2)}{\sigma_{1.34\cdots n}\sigma_{2.34\cdots n}}. \tag{7}$$

Each μ_k is the negative of the ratio of the cofactor of r_{1k} to the cofactor of r_{11} in the determinant obtained by striking out the second row and the second column from R. We shall use the notation $R_{hi\cdot jk}$ to mean the algebraic complement of the second order minor in R, whose complement is obtained by striking out row h and column i and then row j and column k. Then

$$\mu_k = \frac{R_{22\cdot 1k}}{R_{22\cdot 11}}.$$

By argument similar to that used in (3),

$$\Sigma(x_1 - Y_1)Y_2 = 0,$$

or

$$\Sigma x_1 Y_2 = \Sigma Y_1 Y_2.$$

Similarly,

$$\Sigma x_2 Y_1 = \Sigma Y_1 Y_2 .$$

Then the numerator of the right member of (7) becomes, after expanding and collecting terms,

$$r_{12} - \sigma_{Y_1} r_{x_2 Y_1} . \tag{8}$$

Multiplying both sides of (6) by $\frac{x_2}{N}$ and summing over the N individuals, we have,

$$\begin{aligned} \sigma_{Y_1} r_{x_2 Y_1} &= r_{23}\mu_3 + r_{24}\mu_4 + \cdots + r_{2n}\mu_n \\ &= \frac{1}{R_{22\text{-}11}}(r_{23}R_{22\text{-}13} + r_{24}R_{22\text{-}14} + \cdots + r_{2n}R_{22\text{-}1n}) \\ &= r_{12} + \frac{R_{12}}{R_{22\text{-}11}} . \end{aligned} \tag{9}$$

Analogous to (5), we have,

$$\sigma^2_{1.34\cdots n} = \frac{R_{22}}{R_{22\text{-}11}}, \qquad \sigma^2_{2.34\cdots n} = \frac{R_{11}}{R_{11\text{-}22}} . \tag{10}$$

From (8), (9), and (10) the right member of (7) becomes

$$\frac{-R_{12}}{\sqrt{R_{11}R_{22}}} .$$

It is seen that in general

$$r_{jk.12\cdots j-1,j+1,\cdots,k-1,k+1,\cdots n} = \frac{-R_{jk}}{\sqrt{R_{jj}R_{kk}}} .$$

University of Minnesota,

NOTE ON REGRESSION FUNCTIONS IN THE CASE OF THREE SECOND ORDER RANDOM VARIABLES

By Clyde A. Bridger

The study of the correlation of two second-order random variables has received the attention of several authors, among them Yule [1], Charlier [2], Wicksell [3, 4], and Tschuprow [5]. Yule writes of them under the guise of "attributes." The study of three or more second order random variables has lagged behind. In this note we shall examine the regression function of one second order random variable on two others by considering the problem from the point of view of Tschuprow's [6] paper on the correlation of three random variables.

A variable X that takes on m values $x_1, \cdots, x_m$ with corresponding probabilities $p_1, \cdots, p_m$ subject to the condition $\sum_i p_i = 1$ is defined as a random variable of order m. (In particular, if X takes on only two values, x and x' with probabilities p and q, where $p + q = 1$, X is a random variable of second order.) The system of values x and probabilities p constitute the law of distribution of X. In the case of two random variables, X and Y, there exists a joint distribution law, covering all possible combinations of X and Y, together with their associated probabilities $p_{11}, \cdots, p_{mn}$ the joint distribution law contains all of the information regarding the stochastical dependence of X and Y.

The extension to more than two variables is immediate. Let p_{ijk} represent the probability of the simultaneous occurrence of the set of values x_i, y_j, z_k of three random variables X, Y, and Z; p_{ij}, that of the simultaneous occurrence of x_i, y_j together without reference to Z; p_i, that of the occurrence of x_i without reference to Y or Z; etc. Then, we have relationships of the types $\sum_i \sum_j \sum_k p_{ijk} = \sum_i \sum_j p_{ij} = \sum_i p_i = 1$; $\sum_i p_{ijk} = p_{jk}$; $\sum_i \sum_j p_{ijk} = \sum_j p_{jk} = \sum_i p_{ik} = p_k$. Similarly, let $p_{jk}^{(i)}$ be the probability of the simultaneous occurrence of y_j and z_k on the condition that X takes on the value x_i; $p_j^{(i)}$, that of the occurrence of y_j without reference to Z, on the same condition; etc. Then

$$\sum_k p_k^{(i)} = \sum_k p_k^{(ij)} = \sum_j \sum_k p_{jk}^{(i)} = 1; \qquad \sum_j p_{jk}^{(i)} = p_k^{(i)}; \qquad p_i p_j^{(i)} = p_{ij};$$

$$p_{ij} p_k^{(ij)} = p_i p_{jk}^{(i)} = p_i p_j^{(i)} p_k^{(ij)} = p_{ijk}; \qquad \sum_i p_i p_j^{(i)} = p_j; \text{ etc.}$$

Denoting by $E(x)$ or simply Ex the expression "the mean value or mathematical expectation of x," we have $m_{fgh} = EX^f Y^g Z^h = \sum_i \sum_j \sum_k p_{ijk} x_i^f y_j^g z_k^h$. In particular, the mean values of the distributions are given by $m_x = EX$

$= \sum_i p_i x_i$, $m_Y = EY = \sum_j p_j y_j$, $m_Z = EZ = \sum_k p_k y_k$. Then we may write $\mu_{fgh} = E(X - m_X)^f(Y - m_Y)^g(Z - m_Z)^h = Eu^f v^g w^h = \sum_i \sum_j \sum_k p_{ijk}(x_i - m_X)^f (y_j - m_Y)^g (z_k - m_Z)^h$. The quantities μ may be identified as terms in the expression for the moments for the sum of three variables as follows: $E(u + v + w)^n = Eu^n + nEu^{n-1}v + \cdots + kEu^f v^g w^h + \cdots + Ew^n$, where $f + g + h = n$. If $n = 2$, we have the variance of the sum of three variables given by $\mu_{2..} + 2\mu_{11.} + \mu_{.2.} + 2\mu_{.11} + 2\mu_{1.1} + \mu_{..2}$, where the dots in the subscripts indicate variables not considered. Thus $\mu_{2..}$ refers to the second moment of the distribution of the variable X about its mean, m_X, without consideration of the distributions of Y or Z. If every term of the expansion of the n-th moment of the sum of three variables is divided by the quantity $\sqrt{\mu_{2..}{}^f \mu_{.2.}{}^g \mu_{..2}{}^h}$, the expansion takes the "normal form." The type term is $r_{fgh} = \mu_{fgh}/\sqrt{\mu_{2..}{}^f \mu_{.2.}{}^g \mu_{..2}{}^h}$. In the case of one variable, $r_f = \mu_f/\sqrt{\mu_2{}^f}$, so $r_1 = 0$, $r_2 = 1$, $r_3 = \sqrt{\beta_1}$, $r_4 = \beta_2$, etc. In the case of two variables, $r_{1.} = r_{.1} = 0$, $r_{2.} = r_{.2} = 1$, r_{11} = Pearson's product-moment coefficient of correlation, etc. Functions of parameters r will serve to characterize the law of correlation among the variables.

By writing the expressions with superscript (i) to denote that the values of the distributions of Y and Z are those which correspond to a fixed value x_i of the distribution of X, we have $m_Y{}^{(i)} = (EY)^{(i)}$, $m_Z{}^{(i)} = (EZ)^{(i)}$, $\mu_{gh}{}^{(i)} = E(Y - m_Y{}^{(i)})$ times $(Z - m_Z{}^{(i)})$, $r_{gh}{}^{(i)} = \mu_{gh}{}^{(i)}/\sqrt{\mu_{.2.}^{(i)g}\mu_{..2}^{(i)h}}$. (For $g = h = 1$, $r_{gh}{}^{(i)}$ becomes the conditional coefficient of correlation between Y and Z for $X = x_i$.) Thus it follows that we can study the correlation between Y and Z for each value of X separately.

For second order random variables, some changes in notation can be made. Let p_x and $p_{x'}$ be the probabilities corresponding to the values x and x', respectively, of X; p_y and $p_{y'}$ correspond to y and y', respectively; p_z and $p_{z'}$ correspond to z and z', respectively. Also, let p_{xy} represent the probability of the simultaneous occurrence of x and y together without reference to the distribution of Z, etc., and p_{xyz} represent the probability of the simultaneous occurrence of all three values, x, y, z, of their respective distributions, etc. Then, $p_x + p_{x'} = p_y + p_{y'} = p_z + p_{z'} = 1$; $p_{xy} + p_{xy'} = p_x$; $p_{xyz} + p_{xyz'} + p_{xy'z} + p_{xy'z'} = p_x$; etc.

Let us set up a system of normal coordinates in which the values U_i along the U-axis are defined by $U_i = \dfrac{x_i - m_X}{\sqrt{\mu_{2..}}}$, those along the V-axis by $V_j = \dfrac{y_j - m_Y}{\sqrt{\mu_{.2.}}}$, and those along the W-axis by $W_k = \dfrac{z_k - m_Z}{\sqrt{\mu_{..2}}}$. Let $m_Z{}^{(ij)}$ represent the mean of the set of values of the Z distribution which correspond to the fixed pair of values, (x_i, y_j), of the X and Y distributions. Then, in the new coordinate system, the same thing is given by $M_W{}^{(ij)} = \dfrac{m_Z{}^{(ij)} - m_Z}{\sqrt{\mu_{..2}}}$. Now, the series of values $M_W{}^{(ij)}$ obtained by giving i and j different values for the pair (U_i, V_j) determine what is called the regression function of W on U and V (or, in the

original notation, the surface of regression of the distribution of Z on the distributions of X and Y). Similarly, the values of $[M_W^{(ij)}]^{(i)} = \dfrac{m_Z^{(ij)} - m_Z^{(i)}}{\sqrt{\mu_{..2}}}$ obtained by fixing U and varying V in the set (U_i, V_j) determine what is called the conditional line of regression of W on V for a fixed value of U. With these definitions we shall consider the problem of finding a regression function of W on U and V for three second order random variables.

For convenience, write $\delta_{xy} = p_{xy} - p_x p_y$, $\delta_{xz} = p_{xz} - p_x p_z$, $\delta_{yz} = p_{yz} - p_y p_z$, $\alpha_{yz} = p_x p_{xyz} - p_{xy} - p_{xz}$, $\epsilon_z = p_{xyz} - p_{xy} p_z$, $\beta_{yz} = p_{x'} p_{x'yz} - p_{x'y} p_{x'z}$, $\theta_{xyz} = \epsilon_z - p_y \delta_{xz} - p_x \delta_{yz} = \epsilon_y - p_x \delta_{yz} - p_z \delta_{xy} = \epsilon_x - p_y \delta_{xz} - p_z \delta_{xy}$. Direct substitutions into the several formulas developed above then gives us the representative forms to be used in subsequent calculations:

$$x - m_X = p_{x'}(x - x'), \quad x' - m_X = -p_x(x - x').$$

$$m_X = p_x x + p_{x'} x', \quad r_{1..} = 0, \quad r_{2..} = 1, \quad r_{3..} = \frac{p_{x'} - p_x}{\sqrt{p_x p_{x'}}},$$

$$r_{4..} = \frac{1}{p_x p_{x'}} - 3, \quad r_{11.} = \frac{\delta_{xy}}{\sqrt{p_x p_{x'} p_y p_{y'}}}, \quad r_{21.} = r_{11.} r_{3..},$$

$$r_{12.} = r_{11.} r_{.3.}, \quad r_{13.} = r_{11.} r_{.4.}, \quad r_{22.} = r_{11.} r_{3..} r_{.3.} + 1,$$

$$r_{12.} r_{21.} = r_{11.}(r_{22.} - 1), \quad r_{211} = r_{3..} r_{111} + r_{.11},$$

$$r_{121} = r_{.3.} r_{111} + r_{1.1}, \quad r_{112} = r_{..3} r_{111} + r_{11.},$$

$$r_{111} = \frac{\theta_{xyz}}{\sqrt{p_x p_{x'} p_y p_{y'} p_z p_{z'}}}, \quad U_1 = \frac{p_{x'}}{\sqrt{p_x p_{x'}}}, \quad U_2 = \frac{-p_x}{\sqrt{p_x p_{x'}}},$$

$$M_W^{(11.)} = \frac{\epsilon_z}{p_{xy}\sqrt{p_z p_{z'}}}, \quad M_W^{(12.)} = \frac{\delta_{xz} - \epsilon_z}{p_{xy'}\sqrt{p_z p_{z'}}},$$

$$M_W^{(21.)} = \frac{\delta_{yz} - \epsilon_z}{p_{x'y}\sqrt{p_z p_{z'}}}, \quad M_W^{(22.)} = \frac{\epsilon_z - \delta_{xz} - \delta_{yz}}{p_{x'y'}\sqrt{p_z p_{z'}}},$$

$$[M_W^{(11.)}]^{(1..)} = \frac{\alpha_{yz}}{p_{xy}\sqrt{p_{xz} p_{xz'}}}, \quad [M_W^{(21.)}]^{(2..)} = \frac{\beta_{yz}}{p_{x'y}\sqrt{p_{x'z} p_{x'z'}}},$$

$$[M_W^{(12.)}]^{(1..)} = \frac{-\alpha_{yz}}{p_{xy'}\sqrt{p_{xz} p_{xz'}}}, \quad [M_W^{(22.)}]^{(2..)} = \frac{-\beta_{yz}}{p_{x'y'}\sqrt{p_{x'z} p_{x'z'}}}.$$

In the case of correlation of two second order random variables, a linear regression function can always be found [3, 5]. Similarly, the conditional regression functions in the case of three second order random variables can always be taken as linear. If we take as the form of the regression function of W on U and V the form $M_W^{(ij)} = aU_i + bV_j + cU_iV_j + d$, where a, b, c, d are constants to be determined by direct substitution for U_i and V_j from the distributions of X and Y, it is seen that linearity of all total and conditional

regression functions is preserved. By total regression function, we mean the regression of W on U or W on V.

Now consider the problem of finding a, b, c, d. The direct substitution provides us with four linearly dependent equations in four unknowns. Linear combinations reduce the set to three, from which the relationship $d = -cr_{11}.$ is obtained. By building up the various terms in the equations through dividing by the necessary values of p, the parameters r can be made to appear. Further combinations now reduce the set to the following three:

$$\begin{aligned} r_{111} &= ar_{21}. + br_{12}. + c(r_{22}. - r_{11}^2.) \\ r_{.11} &= ar_{11}. + b \quad + cr_{12}. \\ r_{1\cdot 1} &= a \quad + br_{11}. + cr_{21}. \end{aligned}$$

The solution gives

$$a = \frac{r_{1\cdot 1} - r_{11}.r_{.11}}{1 - r_{11}.^2} - \frac{r_{21}. - r_{11}.r_{12}.}{1 - r_{11}.^2} c = a' - a''c$$

$$b = \frac{r_{.11} - r_{11}.r_{1\cdot 1}}{1 - r_{11}.^2} - \frac{r_{12}. - r_{11}.r_{21}.}{1 - r_{11}.^2} c = b' - b''c$$

$$c = (1 - r_{11}.^2)(r_{111} - a'r_{12}. - b'r_{21}.) \div \Delta, \qquad \text{where}$$

$$\Delta = \begin{vmatrix} 1 & r_{11}. & r_{21}. \\ r_{11}. & 1 & r_{12}. \\ r_{21}. & r_{12}. & r_{22}. - r_{11}.^2 \end{vmatrix}$$

The regression function becomes
$M_W^{(ij)} = a'U_i + b'V_j - c(r_{111} + a''U_i + b''V_j - U_iV_j)$. If $c = 0$ the surface is a plane. Examination of the characteristics of r_{111} shows that generally c cannot be zero. The vanishing of c implies that special relations must exist between p_{ijk} and p_{ij}, p_{ik}, p_{jk}.

Two constants of considerable importance in the theory of correlation are the multiple correlation coefficient and the multiple correlation ratio. For the regression of W on U and V, the former is defined as $R_{11}.^2 = a'r_{1\cdot 1} + b'r_{.11}$ and the latter as $\eta_{--2} = \sum_i \sum_j p_{ij}[M_W^{(ij)}]^2$. For planar regression, the difference $\eta_{--2} - R_{11}.^2$ must vanish. For others, the difference takes on values characteristic of the regression function. To find the value it takes for our case, we set up the value of η_{--2} from the regression function just given and subtract $R_{11}.^2$.

By direct substitution, we have $\eta_{--2} - R_{11}.^2 = \sum_i \sum_j p_{ij}(aU_i + bV_j - cU_iV_j - cr_{111})^2 - a'r_{1\cdot 1} - b'r_{.11}$. Since $\sum_i \sum_j p_{ij}U_i^2 = 1$, $\sum_i \sum_j p_{ij}(U_iV_j)^2 = r_{22}.$, etc., we find rather easily that

$$\eta_{--2} - R_{11}.^2 = c^2(r_{22}. - r_{11}.^2) - a''r_{21}. - b''r_{12}..$$

We can also obtain the same value of $\eta_{-2} - R_{11}.^2$ by direct substitution for the four values of $M_W^{(ij)}$ in η_{-2} and subtracting $R_{11}.^2$. To actually obtain this is a long laborious process complicated by the fact that so many alternate forms for the answer are possible, of which only one is comparable with the value previously found. The general procedure is first to set up from the definition the expression $K = p_z p_{z'} \eta_{-2} =$

$$p_{xy}\left(\frac{\epsilon_z}{p_{xy}}\right)^2 + p_{xy'}\left(\frac{\delta_{xz} - \epsilon_z}{p_{xy'}}\right)^2 + p_{x'y}\left(\frac{\delta_{yz} - \epsilon_z}{p_{x'y}}\right)^2 + p_{x'y'}\left(\frac{\epsilon_z - \delta_{xz} - \delta_{yz}}{p_{x'y'}}\right)^2.$$

Then we build up each square by addition and subtraction so that it will contain a θ_{xyz} term. At the close of the process, we convert the whole expression into the parameters r by dividing through by $p_z p_{z'}(p_x p_{x'} p_y p_{y'})^2$ and substituting from the list of representative forms given at the beginning of the paper. A matter of rearrangement now gives the same result as before.

From the symmetry involved, we can say that, in the case of the correlation of three second order random variables, the function representing the regression of one on the other two has an equation in normal coordinates of the form $M_W^{(ij)} = aU_i + bV_j - cU_iV_j - cr_{111}$, where a, b, and c satisfy equations of type

$$r_{111} = ar_{21\cdot} + br_{12\cdot} + c(r_{22\cdot} - r_{11\cdot}^2)$$

$$r_{\cdot 11} = ar_{11\cdot} + b \quad + cr_{12\cdot}$$

$$r_{1\cdot 1} = a \quad + br_{11\cdot} + cr_{21\cdot}$$

State Division of Public Health,
Boise, Idaho.

BIBLIOGRAPHY

[1] G. Udny Yule. *An Introduction to the Theory of Statistics.* London: Charles Griffin & Co., Ltd., 1922. 6th Ed.

[2] C. V. L. Charlier. "Om korrelation mellan eganskaper inom den homograda statistiken." *Svenska Aktuarieföreningens Tidskrift.* Vol. I (1914), pp. 21–35.

[3] S. D. Wicksell. "Some theorems in the theory of probability, with special reference to their importance in the theory of homograde correlation." *Svenska Aktuarieföreningens Tidskrift.* Vol. III (1916), pp. 165–213.

[4] S. D. Wicksell. "On the correlation of acting probabilities." *Skandinavisk Aktuarietidskrift.* Vol. I (1918), pp. 98–135.

[5] A. A. Tschuprow. *Grundbegriffe und Grundprobleme der Korrelationstheorie.* Leipzig: B. G. Teubner, 1925.

[6] A. A. Tschuprow. (Translation into English by L. Isserlis.) "The Mathematical Theory of the Statistical Methods Employed in the Study of Correlation in the Case of Three Variables." *Transactions of the Cambridge Philosophical Society.* Vol. XXIII, no. 12 (1928), pp. 337–382.

THE ANNALS OF MATHEMATICAL STATISTICS

... publication in the ANNALS OF MATHEMATICAL STATISTICS ... Wilks, Fine Hall, Princeton, New Jersey. Manuscripts ... double spaced ... with wide margins, and the original copy ... should be reduced to a minimum and whenever ... at the end of the paper. Formulae ... should be drawn on ... they are to ... typographical difficulties ...

... Fifty reprints without ... furnished at cost ... Single copies ...